INTEREST RATE MODELS WITH NON-GAUSSIAN DRIVEN STOCHASTIC VOLATILITY

By

Jiangchun Bi

Submitted for the Degree of

Doctor of Philosophy

at Heriot-Watt University

on Completion of Research in the

Department of Actuarial Mathematics & Statistics

October 2009.

I hereby declare that the work presented in this thesis was carried out by myself at Heriot-Watt University, Edinburgh, except where due acknowledgement is made, and has not been submitted for any other degree.

Jiangchun Bi (Candidate)

Professor Andrew J. G. Cairns (Supervisor)

Date

Contents

List of Tables

List of Figures

Abstract

In this thesis, we consider some two-factor short rate models that incorporate stochastic volatility with jumps. The motivation for studying such kinds of model is to overcome the shortcomings of diffusion-based stochastic models and to provide a more accurate description of the empirical characteristics of the short rates. In our first model, a jump process for the short-rate volatility is described with jump times generated by a Poisson process and with jump sizes following exponential distribution. Secondly, we extend the volatility model further by taking a superposition of two independent jump processes. We present the corresponding Markov chain Monte Carlo estimation algorithm and provide estimation results of candidate model parameters, latent volatility processes and the jump processes using the 3-month U.S. Treasury Bill rates. Finally, we apply our models to price fixed-income products through Monte Carlo simulation.

Acknowledgements

I would like to express deep and sincere gratitude to my supervisor, Professor Andrew J.G. Cairns for his encouragement and guidance. His wide knowledge and his scientific instinct have been of great value for me. His detailed and constructive comments provide a good basis for the present thesis.

I owe a sincere gratitude to Yi Tian, without her understanding and support it would have been impossible for me to finish this work. With her accompany, the life in Edinburgh is such a pleasant memory. In addition, I have received many encouragement and support from friends during this time, Shumu, Chenming, Ken and Keli.

Finally, my special gratitude is due to my family, my parents and my brother, who provide me such opportunity.

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	Jiangchun Bi		
School/PGI:	School of Mathematical and Computer Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought (Award **and** Subject area)	Doctor of Philosophy

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	

Submission

Submitted By *(name in capitals):*	
Signature of Individual Submitting:	
Date Submitted:	

For Completion in Academic Registry

Received in the Academic Registry by *(name in capitals):*			
Method of Submission *(Handed in to Academic Registry; posted through internal/external mail):*			
*E-thesis Submitted (**mandatory for final theses from January 2009)***			
Signature:		Date:	

Please note this form should bound into the submitted thesis.

Updated February 2008, November 2008, February 2009

Chapter 1

Introduction

1.1 The Interest Rate Model World

One of the most interesting challenges in finance research is to explain the random behaviour of interest rate. As we know, even the movement of riskless interest rate behaves as unpredictably as the asset return rate. However, the evolution of interest rates have different characteristics (non-negative, for example) compared with asset returns. We should make corresponding adjustment when we try to introduce the models used to describe the asset return into interest rate world.

The top left plot in Figure 1.1 presents the evolution of the 3-month U.S. Treasury Bill yield from 1954 to 1997. The short rate climbs to its peak around 1980 and then experiences dramatic falls during the subsequent years. The bottom left plot in Figure 1.1 presents the first difference of the 3-month T-bill yields. There are several points worthy of note. First, the change of short rate is not stable, high changes are followed by lower ones or jump to higher levels for some periods. Second, the changes are sensitive to the level of the short rate. When the rates are high, the changes for this period are also bigger than any other time. Let's try one naive model to capture the dynamics of short rate. The bottom right plot in Figure 1.1 presents the Brownian motion with the same sample mean and variance as the first difference of short rate. Obviously, these two processes are different. Neither the shapes nor the value ranges of them are similar. The naive Brownian motion is a

poor candidate for the short rate. The evolution of the short rate generated from such a model will be like the top right plot in Figure 1.1.

Pioneering interest rate models focus on the instantaneous short-rate process which is continuous and latent, such as the Vasicek (1977) model, the Dothan (1978) model, and the Cox *et al.* (1980) (CIR) model. Although analytical pricing formula for some simple products can be deduced from such models, their ability in explaining the historical data is quite poor. The calculations based on such model are quite complicated and time-consuming when dealing with exotic products. As the interest rate market developed, more and more sophisticated models have been developed to fit the needs of more accurate, reasonable, and explainable pricing for interest rate related derivatives during the last thirty years. Not only more factors but also other interest rates, such as instantaneous forward rates and LIBOR have been considered as the model objects. For example, Richards (1978) considers both the short rate and the inflation rate as the state variables, and Brace *et al.* (1997) considers the full set of LIBOR rates as the state variables.

In this thesis, we introduce a random volatility model for interest rates which is driven by a jump process. Such processes have been applied before in the construction of asset price models but have not previously been used to describe the volatility of interest rate. We compare various aspects of the performance of the new model against several alternative models, including a single factor model, the Chan, Karoli, Longstaff and Sanders (CKLS) model (Chan *et al.* (1992)), a stochastic volatility model driven by a Gaussian process and a variation of the new model. In this variation model, the driven process is a superposition of jumps. We compare their ability to capture the behaviour of empirical data by residual analysis. We obtain outstanding improvements by introducing a jump process in the volatility specification. We also compare the capability of the models to price different interest rate derivative products. Our estimation algorithm for the models is partly inspired by the work of Papaspiliopoulos (2003), although we have had to adapt his procedure to generate the jump process and update the parameters of that process.

1.2 Thesis Outline

In Chapter 2, we discuss short rate models with stochastic volatility in particular. The SV models we will examine are all driven by Lévy processes. One of them is based on the Wiener process and others are linked to a Compound Poisson process. The latter models introduce jumps into the volatility and it will be demonstrated that these provide much better candidate processes for describing the historical behavior of the short rate.

In Chapter 3, we describe the MCMC algorithms used for parameter estimation for the CKLS model, the SVG model, the SVJ model and the SVS model respectively. We also provide simulation tests to assess the efficiency of each algorithm by simulating the short rate from the corresponding model given predetermined parameter values.

We apply these algorithms in Chapter 4 to the 3-month U.S. T-bill rate from 1954-1997. The estimation results are accompanied by statistical tests for the normality and independence of the model residuals. We also examine the Bayes statistic which is the appropriate model selecting benchmark since it considers about the extra parameters (factors) introduced by the SV models. We examine the empirical volatility from the SV models towards the end of this chapter.

In Chapter 5, we apply these models in pricing fixed-income products, including zero-coupon bonds, bond options and caps by Monte Carlo simulation. We will use the estimation results from Chapter 4 in simulation which is in the canonical measure. We first figure out the pricing measure for each model. For the zero-coupon bond, we investigate of the yield curves generated under each model. For both options and caps, we check the implied volatilities calculated from each model. Finally in Chapter 6, we give conclusions and suggestions for further research.

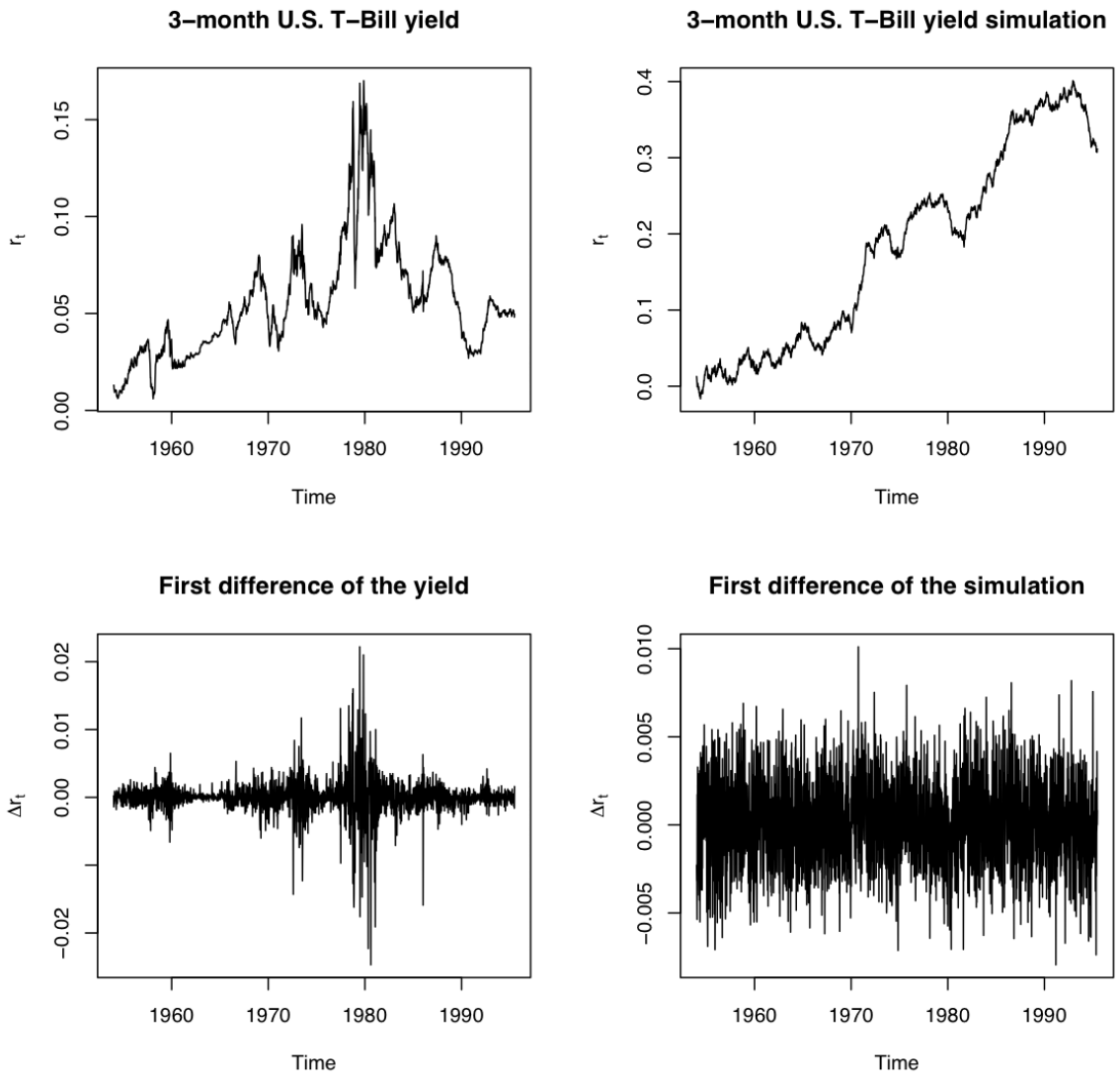

Figure 1.1: Historical data (3-month U.S. T-Bill) compare with the simulation (A Gaussian process with same mean and variance).

In the following sections, we first review some basic definitions of interest rates, and the common fixed-income products. We describe basic pricing theory and some well known interest-rate models. Three estimation methods are also presented in this chapter. The remaining chapters consider a number of existing and new models for the short rate with stochastic volatility. The models are including the CKLS model, a stochastic volatility model driven by Gaussian process (SVG), a stochastic volatility model driven by jump process (SVJ) and a stochastic volatility model driven by a superposition of jump processes (SVS).

1.3 Basic Definitions of Interest Rate

1.3.1 Money-Market Account and Instantaneous Spot Rate

Interest rates are used to measure the time value of money. One unit of currency today is not equal to one unit of currency a year later. Interest is the compensation for the delayed payment. Let $B(0,t)$ be the accumulated value of the *money-market account* of 1 unit currency deposited at time 0 compounded continuously to time t. $B(0,t)$ is:

$$B(0,t) = \exp\left\{\int_0^t r_s ds\right\}, \tag{1.1}$$

and it satisfies the stochastic differential equation:

$$dB(0,t) = r_t B(0,t) dt, \tag{1.2}$$

where r_t is the *instantaneous spot rate*. $B(0,t)$ is a random variable if r_t is a random process, or a deterministic function of time if $r_t = r$ a constant parameter.

1.3.2 Bond and Spot Rate

A *zero-coupon bond* is a contract that guarantees its holder a single fixed payout at maturity, and issued at discount. Let $P(t,T)$ be the price at time t of a zero-coupon bond with a final payment of 1 unit of currency at maturity T. The *spot*

rate, $R(t,T)$ is defined as a continuously compounded constant yield at time t for the zero-coupon bond that matures at T.

$$R(t,T) := -\frac{\log P(t,T)}{T-t}, \tag{1.3}$$

for $T > t$. The *instantaneous spot rate*, r_t, is the limit of $R(t,T)$ as T converges to t,

$$r_t = \lim_{T \to t} R(t,T). \tag{1.4}$$

Given a set of zero-coupon bond prices, $\{P(t,T_i)\}_i$, the *yield curve* can be composed by plotting the spot rate, $R(t,T_i)$ against the maturity time T_i, which is also known as the *term structure of interest rate*. It is a useful tool not only for bond pricing through Equation (1.3), but also to predict future changes of the interest rate.

1.3.3 Forward Rate and Instantaneous Forward Rate

The *forward rate*, $F_t(T,S)$ is the constant continuously compounding rate applied over the future period (T,S) but fixed at time t. It is defined by

$$F_t(T,S) := -\frac{1}{\tau} \log \frac{P(t,S)}{P(t,T)}, \tag{1.5}$$

where $\tau = S - T$, usually called the *tenor*, and $S \geq T$. The *instantaneous forward rate* $f_t(T)$ is the limit of forward rate, defined by:

$$f_t(T) := \lim_{\tau \to 0} F_t(T, T+\tau). \tag{1.6}$$

It is the interest rate known at time t, for instantaneous borrowing at time T. Given Equation (1.5), the price of a zero coupon bond $P(t,T)$ can also be written as:

$$P(t,T) = \exp\left\{ -\int_t^T f_t(u)du \right\}. \tag{1.7}$$

Hence, with Equation (1.3), the relation between $R(t,T)$ and $f_t(T)$ is:

$$f_t(T) = R(t,T) + (T-t)\frac{\partial}{\partial T} R(t,T). \tag{1.8}$$

Therefore, $R(t,T) = F_t(t,T)$ and $r_t = f_t(t)$.

1.3.4 London Interbank Offered Rate

London interbank offered rate (**LIBOR**) is the interest rate at which banks are willing to lend to each other for a period. LIBOR is an annualized simple composed interest rate paid at the end of the defined period, therefore we have the following equation:

$$P(t, t+\tau)(1 + L(t, t+\tau)\tau) = 1,$$

where τ known as the *year fraction*, is the length in years of the period $(t, t+\tau)$. Here $L(t, t+\tau)$ is the spot LIBOR rate at t.

A forward LIBOR rate, $L(t, T, T+\tau)$ is the simple composed forward rate at t applying over the period $[T, T+\tau]$, defined by:

$$L(t, T, T+\tau) = \frac{1}{\tau}\left(\frac{P(t,T)}{P(t,T+\tau)} - 1\right). \tag{1.9}$$

The relationship between spot LIBOR and forward LIBOR is $L(T, T+\tau) = L(T, T, T+\tau)$. For simplicity, we denote $L(t, T, T+\tau) = L_t(T, \tau)$ in later sections.

1.4 Interest Rate Derivatives

A bond is one of the most common interest rate derivatives but some other basic products will be reviewed in this section. They provide the foundation for most exotic derivatives. We will also examine the pricing ability of our candidate models through these products in later chapter.

1.4.1 Swaps

A vanilla interest rate *swap* is an agreement between party A and B, to exchange a series of interest payments based on the same principal. At the specified dates, party A will pay to party B interest at a predetermined fixed rate, K, in exchange

for a simultaneous interest payment from party B corresponding to a floating rate on the same principal. The fixed rate K is known as the *swap rate* which is usually chosen to make zero initial premium to either party. The swap contract for the party that makes fixed payments is called as a *payer swap*, and for those receiving fixed payments is called a *receiver swap*.

Usually, LIBOR is used for the floating rate of a swap. Consider a swap with N payment dates, say $T_1, ..., T_N$ and let $T_0 < T_1$ be the current time. The net income for the payer party in the period (T_{i-1}, T_i) is:

$$V_p(T_{i-1}, T_i) = \alpha_{i-1}(L(T_{i-1}, T_i) - K), \tag{1.10}$$

where the payment is made at time T_i, α_{i-1} is the year fraction between T_{i-1} and T_i, and $L(T_{i-1}, T_i)$ is the spot LIBOR rate. Such payment is also known as the *swaplet*.

From Equation (1.10), a payer swap contract can be decomposed into a portfolio of two kinds of coupon bonds: with a long position in a floating-rate coupon bond and shorting a fixed-rate coupon bond. The floating-rate coupon bond is always equal to its par value since the same floating rate is used as the discount rate. The present value of the fix-rate coupon bond is a sum of zero-coupon bond with different final payments and maturities. Therefore, the present value of a payer swap at T_i will be:

$$V_p(T_i) = 1 - \sum_{j=i+1}^{N} K\alpha_{j-1}P(T_i, T_j) - P(T_i, T_N). \tag{1.11}$$

Similarly, the value of receiver swap V_r is equal to a portfolio with a long position in a fixed-rate coupon bond while shorting a floating one. Its present value at T_i, V_r is equal to $-V_p$. Since the swap rate is selected to make the initial value of the contract equal to zero, $V_p = V_r = 0$, we obtain:

$$K = \frac{1 - P(T_0, T_N)}{\sum_{i=1}^{N} \alpha_{i-1}P(T_0, T_i)}. \tag{1.12}$$

1.4.2 Bond Options

Bond options are the financial instruments which gives the holders the right but not the obligation to buy(call option) or sell (put option) the specified bond at the predetermined price (strike price) on a particular date (European options) or over a particular period (American options).

For a European call option on a zero-coupon bond maturing at T, the payoff at expiring date S with strike price K is then:

$$\max(P(S,T) - K, 0),$$

where $P(S,T)$ is the price of the underlying bond at S, $S \leq T$. With an identical setting, the payoff of a European put option is:

$$\max(K - P(S,T), 0).$$

1.4.3 Caps and Floors

Interest rate caps are options under which the holder has the right to be compensated by the writer if a reference interest rate is higher than a predetermined strike rate (*cap rate*). Interest rate floors are similar contracts with the payments occurring when the reference rate is lower than the strike rate (*floor rate*).

LIBOR is usually selected as the reference rate. Let $T_1, ..., T_N$ be the payment dates. At time T_i, the ith payoff for a cap is:

$$\alpha_{i-1} \max(L(T_{i-1}, T_i) - r_c, 0),$$

where α_{i-1} is the year fraction for the period (T_{i-1}, T_i), and r_c is the cap rate. The option that expires at time T_i with that payoff is called as the *caplet*. It is a European call option on the LIBOR rate. The present value of that caplet at T_{i-1} is then:

$$\frac{1}{1 + \alpha_{i-1} L(T_{i-1}, T_i)} \alpha_{i-1} \max(L(T_{i-1}, T_i) - r_c, 0)$$

which can be written as:

$$\begin{aligned} & \max\left(1 - \frac{1+\alpha_{i-1}r_c}{1+\alpha_{i-1}L(T_{i-1},T_i)}, 0\right) \\ = \; & \max\left(1-(1+\alpha_{i-1}r_c)P(T_{i-1},T_i), 0\right). \end{aligned}$$

Therefore, the caplet at T_{i-1} has the same value as a put option on a zero-coupon bond with strike price equal to $\frac{1}{1+\alpha_{i-1}r_c}$ exercised at T_{i-1}. A cap is a portfolio of call options on the LIBOR rates. The stochastic discounted payoff to time $t < T_0$ is given by:

$$\sum_{i=1}^{N} B(t,T_i)^{-1}\alpha_{i-1}\max(L(T_{i-1},T_i) - r_c, 0), \tag{1.13}$$

or put options on the zero-coupon bonds:

$$\sum_{i=1}^{N} B(t,T_{i-1})^{-1}\max\left(1-(1+\alpha_{i-1}r_c)P(T_{i-1},T_i), 0\right). \tag{1.14}$$

Similarly, at time T_i, the ith payoff for a floor with strike rate r_f, the *floorlet* is:

$$\alpha_{i-1}\max(r_f - L(T_{i-1},T_i), 0)$$

which is an European put option on the LIBOR rate or a put option on a zero-coupon bond.

1.5 Pricing Theory

In an efficient market, the price of a commodity is determined by its supply and demand. For a financial instrument, the relationship between supply and demand is driven, at least in part, by the arbitrage free principle. An arbitrage opportunity exists, if there is an instrument with zero (or negative) price currently but with nonnegative payoffs almost surely in the future. In an efficient market, such opportunities will be discovered quickly. Increasing demand and decreasing supply will then drive the price of that instrument to the balanced *arbitrage free* level. In this

section, we will review two general pricing methodologies for securities based on that principle: the *Martingale* method and the *PDE* method.

1.5.1 Probability Space

The *sample space* Ω is a measurable set containing all possible outcomes, the *sample points*. $\mathcal{F}$ is a *σ-algebra* of subsets of Ω. $\mathbb{P}$ is the *probability measure* on $(\Omega, \mathcal{F})$ defined as a function:

$$\mathbb{P} : \mathcal{F} \to [0, 1],$$

which also has the following properties

- $\mathbb{P}(\Omega) = 1$;
- for any sequence of disjoint sets $\mathcal{A}_n \in \mathcal{F}$,

$$\mathbb{P}(\bigcup_{n\geq 1} \mathcal{A}_n) = \sum_{n\geq 1} \mathbb{P}(\mathcal{A}_n).$$

$(\Omega, \mathcal{F}, \mathbb{P})$ is called the *probability space* or the *probability triple*. A *filtration*, $F = \{\mathcal{F}_t, t \in (0, T)\}$, is a family of σ-algebras where $\mathcal{F}_t$ represents the information available up to and including time t. $\mathcal{F}_t \subset \mathcal{F}_s$ if $t \leq s$.

A random process $\{X_t\}$ is a *martingale* if X satisfies the following three conditions:

1. X_t is adapted to $\mathcal{F}_t$;
2. $E(|X_t|)$ is finite for any $t \in [0, T]$;
3. $E(X_S|\mathcal{F}_t) = X_t$, for $t < S \leq T$.

A probability measure $\mathbb{Q}$ is an *equivalent martingale measure* to $\mathbb{P}$ if:

1. $\mathbb{Q}$ is equivalent to $\mathbb{P}$ (both measures have the same null set);
2. a positive random process X_t is a martingale under $\mathbb{Q}$.

We assume that $P(t,T)$, the price of a zero-coupon bond, depends on a random process X_t over the period $[t,T]$: $P(t,T) = P(t,T,X_s), s \in [t,T]$. Here we will start with a random interest rate process, $X_t = r_t$. The dynamics of the short rate are given by:

$$dr_t = a(t,r_t)dt + b(t,r_t)dW_t, \tag{1.15}$$

where W_t is a standard Brownian motion under the canonical measure (the real world measure) $\mathbb{P}$. Given these settings, we will apply the following two methodologies to figure out the arbitrage-free price for the bond.

1.5.2 Equivalent Martingale Measure

Let S_i $(i = 0, 1, ..., n)$ be the current asset value under the canonical measure $\mathbb{P}$. The relative value of S_i with respect to S_j is $S_i^{'} = S_i/S_j$ $(i, j = 0, 1, ..., n)$ where $S_j > 0$ is called the *numeraire*. Given a numeraire S_0, the fundamental idea of the equivalent martingale measure method is to find a measure $\mathbb{Q}$ equivalent to the canonical measure $\mathbb{P}$, under which the relative value of every financial instrument, $S_i^{'}$, is a martingale. Then the price of this instrument would be a function of its expectation under the new measure $\mathbb{Q}$. The following theorem will help us to find the new measure.

Theorem 1.1 ***Girsanov's theorem****:(Cameron-Martin-Girsanov) Let $W(t)$ be a standard Brownian motion under a measure $\mathbb{P}$, and $\gamma(t)$ be a stochastic process satisfying the Novikov's condition*

$$\int_0^t \gamma(s)^2 ds < \infty$$

almost surely. Then there exists an equivalent measure $\mathbb{P}^*$ with the Radon-Nikodym derivative which is defined by $\rho(t) = \frac{d\mathbb{P}^*}{d\mathbb{P}}$ with the following form:

$$\rho(t) = \exp\left(-\int_0^t \gamma(s)dW(s) - \frac{1}{2}\int_0^t \gamma(s)^2 ds\right).$$

Under measure $\mathbb{P}^*$, $W^*(t) = W(t) + \int_0^t \gamma(s)ds$ is a Brownian motion.

Theoretically, any positive financial instruments without interim payments can be chosen as the numeraires. One of the well-known numeraires is the money-market account $B(0,t)$ (Equation (1.1)) and the corresponding measure is called the *risk neutral measure*, introduced by Harrison & Kreps (1979).

Under the risk neutral measure $\mathbb{Q}$, the relative price of a zero-coupon bond $P'(t,T) = \frac{P(t,T)}{B(0,t)}$ is a martingale:

$$P'(t,T) = E_{\mathbb{Q}}\left(P'(T,T) \mid \mathcal{F}_t\right).$$

We can get the following equation:

$$P(t,T) = E_{\mathbb{Q}}\left(\exp\left(-\int_t^T r_s ds\right)\middle|\, \mathcal{F}_t\right),$$

since $P(T,T) = 1$ and $B(0,t)$ is adapted to $\mathcal{F}_t$. For r_t with the SDE defined by Equation (1.15) under $\mathbb{P}$-measure, the SDE under $\mathbb{Q}$-measure will be:

$$dr_t = (a(t,r_t) - \lambda(t)b(t,r_t))dt + b(t,r_t)dW_t^{\mathbb{Q}}$$

where, at time t, $W_t^{\mathbb{Q}}$ is a standard Brownian motion under $\mathbb{Q}$, $\lambda(t)$ is the *market price of risk* by Girsanov's theorem setting $\gamma(t) = \lambda(t)$.

Similarly, the value of a derivative $H(t,T)$ with $\mathcal{F}_t$-measurable final payoff $H(T,T)$ at the expiry date T is determined by the following equation:

$$H(t,T) = E_{\mathbb{Q}}\left(\exp\left(-\int_t^T r_s ds\right) H(T,T)\middle|\, \mathcal{F}_t\right). \tag{1.16}$$

Therefore, the joint distribution of the random payoff and the interest rate process is required for pricing this derivative under the $\mathbb{Q}$ measure. Such complexity can be avoided by choosing a different numeraire. Instead of $B(0,t)$, if we choose, $P(t,T)$, the zero coupon bond as the numeraire, Equation (1.16) will be:

$$H(t,T) = P(t,T)E_{\mathbb{Q}^T}\left(H(T,T)|\, \mathcal{F}_t\right), \tag{1.17}$$

where $\mathbb{Q}^T$ is the corresponding equivalent martingale measure, the *T-measure*.

1.5.3 Partial Differential Equation

According to the arbitrage free principle, the yield of a risk free portfolio is equal to the short rate r_t. We can construct such a riskless portfolio from two bonds with different maturities T_1 and T_2 in the setting in section 1.5.1 where $P(t,T)$ depends on one random factor, the short rate r_t. With boundary condition $P(T,T) = 1$, it can be proven (Vasicek (1977)) that the PDE of zero-coupon bond is:

$$\frac{\partial P}{\partial t} + (a - b\lambda)\frac{\partial P}{\partial r} + \frac{1}{2}b^2\frac{\partial^2 P}{\partial r^2} - rP = 0, \tag{1.18}$$

where the dynamic function of r is defined by Equation (1.15) and λ is a deterministic variable. Generally, with the different boundary conditions at T, we will obtain the following PDE for any other interest derivatives:

$$\frac{\partial V}{\partial t} + (a - b\lambda)\frac{\partial V}{\partial r} + \frac{1}{2}b^2\frac{\partial^2 V}{\partial r^2} - rV = 0, \tag{1.19}$$

where V is the value of such instrument which is a function of both time and the short rate.

Applying the Feynman-Kac formula (Pelsser (2000)), the solution for PDE (1.18) with the boundary condition $P(T,T) = 1$ will be:

$$P(t,T) = E_{\mathbb{Q}}\left(\exp\left(-\int_t^T r_s ds\right) \mid \mathcal{F}_t\right),$$

and for PDE (1.19) with boundary condition $V(T)$

$$V(t) = E_{\mathbb{Q}}\left(\exp\left(-\int_t^T r_s ds\right) V(T) \mid \mathcal{F}_t\right),$$

under the risk neutral measure $\mathbb{Q}$.

Although the same result can be obtained by either method, the PDE approach offers us a numerical way to calculate the price of interest rate instruments when the analytical formula of these solutions are not available.

1.6 Interest Rate Models

Interest rate models can be classified as different categories according to different criteria. Models are classified as the *affine* models or the *non-affine* models by whether they can provide an affine form bond pricing formulae or not. According to the number of underlying random factors, models are named as the *one-factor* models, the *two-factor* models and so on. Models can also be classified by the interest rates which they focus on: *short rate* models, *forward rate* models, and *market rate* models, for example.

In this section, we will take a brief review of interest rate models by the last kind of classification: models named after the underlying interest rates. Generally, *short rate* models focus on the dynamics of short rate r_t; *HJM* models start from instantaneous forward rate $f(t,T)$; and *market rate* models study the observable market rate (LIBOR, for example) directly. In following sections, we will only discuss these kinds of models with one random factor. A review of multi-factor models can be found in Chapter 2.

1.6.1 Short Rate Models

The CKLS model (Chan *et al.* (1992)) is a general single factor short rate model within which many models are nested. The general form of CKLS model is:

$$dr_t = (\alpha + \beta r_t)dt + \sigma_r r_t^{\gamma} dW_t, \tag{1.20}$$

where W_t is a standard Brownian motion under the risk neutral measure $\mathbb{Q}$, and β presents the rate of r_t conversing to its local long term mean $-\frac{\alpha}{\beta}$; both σ_r and r_t^{γ} compose the local stochastic volatility of r_t. The CKLS model is a time-homogeneous model where all the parameters are constant variables.

Table 1.1 lists the models generated from the CKLS model. Chan *et al.* compared these models using the 3-month US T-Bill rate (monthly) covering the period from 1954 to 1980. They found that the model with volatility highly correlated to its state

Table 1.1: Short rate models nested in CKLS model

Model	α	β	γ
Merton	-	0	0
Vasicek	-	-	0
Cox-Ingersoll-Ross	-	-	0.5
Dothan	0	-	1
Brennan-Schwartz	-	-	1

values ($\gamma > 1$) explains the historical observations better than others. Furthermore there are no suggestions to reject the mean reverting setting. They estimate the CKLS model by the generalized method of moments (GMM, section 1.7.1) and the estimate of γ is around 1.5.

As the parameter γ increases from 0 to higher values, the statistical properties of the underlying instantaneous short rate will be changed significantly. First, the stationary distribution of r_t is changing from the normal distribution with $\gamma = 0$ to fat-tailed distributions, for example, the non-central Chi-square distribution for the CIR model with $\gamma = 0.5$. Additionally $\gamma > 0.5$ prevents r_t from hitting zero or becoming negative. The affine form pricing formula could be obtained with $\gamma = 0$ or $\gamma = 0.5$. Although an analytical formula for pricing is still available when $\gamma = 1$ (the Dothan model), the complexity of the formula makes it unsuitable for directly computation. Fortunately, we can apply numerical methods for calculation when works on the models without the analytical pricing formula.

1.6.2 Heath-Jarrow-Morton Framework

In this section, we will present the framework of Heath *et al.* (1992). Instead of modeling the short rate, they began with the instantaneous forward rate $f(t, T)$ directly. With the whole forward rate curve as the input, the theoretical price from the HJM model matches the market price at the calibration date, say t.

Heath *et al.* assume that $f(t, T)$ is an Itô process following the stochastic differential

equation:

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW_t$$

under the $\mathbb{P}$-measure and parameters $\alpha(t,T)$ and $\sigma(t,T)$ are adapted to the filtration $\mathcal{F}_t$. They also prove that under the risk neutral measure $\mathbb{Q}$, the drift term $\alpha(t,T)$ is restricted to:

$$\alpha(t,T) = \sigma(t,T)\left(\int_t^T \sigma(t,s)ds + \lambda(t)\right),$$

where $\lambda(t)$ is the market price of risk defined as before. Then under the $\mathbb{Q}$-measure, the dynamics of $f(t,T)$ is:

$$df(t,T) = \sigma(t,T)\int_t^T \sigma(t,s)dsdt + \sigma(t,T)d\tilde{W}_t,$$

where $\tilde{W}_t$ is a standard Brownian motion under the $\mathbb{Q}$-measure.

Recall that $r(t) = f(t,t)$, the spot rate follows the underlying stochastic process in HJM under the $\mathbb{Q}$-measure:

$$r(t) = f(0,t) + \int_0^t \sigma(u,t)\int_u^t \sigma(u,s)dsdu + \int_0^t \sigma(u,t)d\tilde{W}_u.$$

1.6.3 Market Models

Directly specifying the dynamics of market rates, like LIBOR and swap rates, we can deduce the valuation formulae for the caps, floors and swaptions more conveniently than the short rate models and forward rate models. First introduced by Miltersen *et al.* (1997), as well as Brace *et al.* (1997) the forward LIBOR, $L_t(T,\tau)$, are assumed to be lognormal random processes under the risk neutral measure $\mathbb{Q}$:

$$dL_t(T,\tau) = L_t(T,\tau)\alpha_{T+\tau}(t)dt + L_t(T,\tau)\sigma_t(T+\tau)d\tilde{W}_t,$$

where $\tilde{W}_t$ is a standard Brownian motion under the $\mathbb{Q}$-measure.

From Equation (1.9), $L_t(T,\tau)P(t,T+\tau)$ is the value of a bond portfolio $(P(t,T) - P(t,T+\tau))/\tau$ which is a tradable asset. If we select the zero-coupon bond $P(t,T+\tau)$ as numeraire, $L_t(T,\tau)$ will be a martingale under the *forward measure* $\mathbb{Q}^{T+\tau}$. The stochastic differential equation of $L_t(T,\tau)$ under the $\mathbb{Q}^{T+\tau}$-measure is:

$$dL_t(T,\tau) = L_t(T,\tau)\sigma_t(T+\tau)dW_t^{T+\tau},$$

where $W_t^{T+\tau}$ is a standard Brownian motion under the $\mathbb{Q}^{T+\tau}$-measure.

It is easily proved that the caplet pricing formulae for this model is:

$$C_{T+\tau}(t) = P(t,T+\tau)(L_t(T,\tau)\Phi(d_1) - k\Phi(d_2)),$$

where k is the strike rate, $C_{T+\tau}(t)$ the caplet value at t for the payments occurring at T, $\Phi()$ is the standard Normal density distribution,

$$d_1 = \frac{\log(L_t(T,\tau)/k) + \frac{1}{2}\sigma_c^2}{\sigma_c}, d_2 = d_1 - \sigma_c$$

and $\sigma_c = \int_t^T \sigma_s^2(T+\tau)ds$.

1.7 Estimation Methodology

In this section, we will give a brief review of some commonly used methods of estimating the parameters of interest rate models, the generalized method of moments (GMM), the maximum likelihood method (MLE), and the Markov chain Monte Carlo method (MCMC).

1.7.1 Generalized Method of Moments

The idea of the *generalized method of moments* (GMM) (Hansen (1982)) is to find the parameter values which minimize the differences between a set of empirical moments and the corresponding theoretical ones. The GMM estimator $\hat{\theta}$ given the sample $\{x_i, i = 1, ..., N\}$ is obtained by solving the underlying equation:

$$\hat{\theta} = \arg\min V^T(\theta)MV(\theta),$$

where M is the weighting matrix obtained from the sample; and $V(\theta)$ is the vector of the differences of moments where the jth element is defined as:

$$V_j(\theta) = \frac{1}{N}\sum_{i=1}^{N} f_j(x_i) - E(f_j(x)|\theta),$$

where $f_j()$ is a function on both samples and parameters. There are many candidates for the selection of $f()$, for example, mean, variance, and other moments. More attention should be paid when choosing $f()$, for poor candidates can lead to biased estimators. Chan *et al.* (1992) apply this method to estimate the CKLS model where they choose $f()$ as a function of the residuals. More details about this approach can be found in James & Webber (2000).

1.7.2 Maximum Likelihood Method

The MLE method begins with the mathematical expression known as the *likelihood function* of the sample. Given $f(x)$, the probability density function of the sample $\{x_i\}$, the likelihood function is defined as:

$$L(\theta) = f(x_1, x_2, ..., x_n|\theta).$$

It can be written as:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta),$$

where the observations are assumed to be independent. The maximum likelihood estimator is the value of θ which would maximize the likelihood function given the sample:

$$\hat{\theta} = \arg\max L(\theta),$$

or equally the log-likelihood function, $l(\theta)$:

$$\begin{aligned}\hat{\theta} &= \arg\max l(\theta) \\ &= \arg\max \sum_{i=1}^{n} \ln f(x_i|\theta).\end{aligned}$$

Estimation are obtained by differentiating L or l with respect to θ. Full description and examples can be found in James & Webber (2000).

These two methods are convenient for some single factor models, however they cannot deal with the model with latent variables. Other methods, like Kalman Filter and methods based on Bayesian theory are used to incorporate such model.

1.7.3 Markov Chain Monte Carlo Method

In both the GMM and MLE methods, the unknown parameters are treated as constants and estimates are calculated directly based the statistical properties of the sample X. Using Bayesian methods by contrast, the parameters θ are treated as random variables. We can generate the sample of θ from its posterior distribution $\pi(\theta|X)$ which is defined as:

$$\pi(\theta|X) = \frac{L(\theta)p(\theta)}{f(X)},$$

where $L(\theta)$ is the likelihood function, $p(\theta)$ is the prior distribution of θ and $f(X)$ is the distribution of X. The estimation of θ is obtained from the sample average:

$$\hat{\theta} = \frac{1}{N-n}\sum_{t=n+1}^{N}\theta^{(t)},$$

where N the length of sample, n, known as the *burn in*, is the number of first sample values ignored as unstable variables and $\theta^{(t)}$ is the tth iteration of the simulation.

The *Monte Carlo Markov Chain* (MCMC) method is designed to draw the unknown parameter sample from the posterior distribution when the analytical form of the distribution is unknown. The algorithm is to generate the sample by constructing Markov chain that has the desired posterior distribution as its stationary distribution. The Metropolis-Hastings (MH) (Metropolis *et al.* (1953), Hasting (1970)) and the Gibbs (Geman & Geman (1984)) samplers are mainly used to build such Markov chains.

The Metropolis-Hastings Sampler

Let $\theta := \{\theta_1, ..., \theta_k\}$, $k \geq 1$ be the k-dimensional vector of parameters, $\theta_{-i} := \{\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_k\}$, be the vector comprising all elements except θ_i.

To simulate the $t+1$th θ_i , we first generate a candidate ξ from the proposal distribution $q(\xi|\theta^{(t)})$. $\theta_i^{(t+1)}$ will either move to the candidate with the acceptance probability $\alpha(\theta_i^{(t)}, \xi)$ which is a measurable function defined by Equation (1.21), or it stays with the previous value $\theta_i^{(t)}$ with the probability $1 - \alpha(\theta_i^{(t)}, \xi)$. Generally speaking, the proposal density $q(\xi|\theta^{(t)})$ can be any form but the convergence rate will be different. More details about the selection of proposal distribution can be found in Gilks *et al.* (1996) . The acceptance probability is defined as:

$$\alpha(\theta_i^{(t)}, \xi) := \min\left(1, \frac{\pi(\xi|\theta_{-i}^{(t)})q(\theta^{(t)}|\xi)}{\pi(\theta_i^{(t)}|\theta_{-i}^{(t)})q(\xi|\theta^{(t)})}\right), \tag{1.21}$$

where $\pi(\theta_i^{(t)}|\theta_{-i}^{(t)})$ is the full conditional distribution for $\theta_i^{(t)}$ and $\theta_{-i}^{(t)} := \{\theta_1^{(t+1)}, ..., \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, ..., \theta_k^{(t)}\}$. The distribution is derived from the joint distribution for vector θ at the t-th interation:

$$\pi(\theta_i^{(t)}|\theta_{-i}^{(t)}) = \frac{\pi(\theta_i^{(t)}, \theta_{-i}^{(t)})}{\int \pi(\theta_i^{(t)}, \theta_{-i}^{(t)}) d\theta_i^{(t)}}, \tag{1.22}$$

where $\pi(\theta_i^{(t)}, \theta_{-i}^{(t)})$ is the joint distribution.

The algorithm of MH is described as below:

`Initialize` $\theta^{(0)}$

`For` t `from` 0 `to N; with increment` 1

{

  `For` i `from` 1 `to k; with increment` 1

  {

    `Sample a candidate` Y `from` $q(.|\theta_{-i}^{(t)})$

    `Sample a random variable` U `uniformly from` $(0, 1)$

    `If` $U \leq \alpha(\theta_i^{(t)}, \xi)$

      `Set` $\theta_i^{(t+1)} = \xi$

Otherwise

Set $\theta_i^{(t+1)} = \theta_i^{(t)}$

}

}

The Gibbs Sampler

The Gibbs sampler generate the candidates from the corresponding full conditional distributions. Since the acceptance probabilities are equal to one, the candidates will be always accepted as the new values. The algorithm is described as below:

Initialize $\theta^{(0)}$

For t starts from 0 to N; with increment 1

{

For i starts from 1 to k; with increment 1

{

Sample a candidate ξ from $\pi(.|\theta_{-i}^{(t)})$

Set $\theta_i^{(t+1)} = \xi$

}

}

The Gibbs sampler is a special case of the MH algorithm. In most cases, both methods are applied in the MCMC algorithm.

Chapter 2

The Stochastic Volatility Models

2.1 Introduction

We have given a brief review about some one-factor models in Chapter 1. Although the one-factor models are more analytically tractable and easier to understand, there are several evidences indicating that further factors are needed to describe the evolution of the interest rate process. Firstly, the spot rates with different maturities from one-factor models are perfectly correlated. It means that the yield curve generated by such models will move in the same direction when a shock to the interest rate happens in the market. In contrast, observations from market have rejected this conclusion and indicated that more factors are required to extend the correlation structures, (Barndorff-Neilsen (2001)). Secondly, the shapes of the yield curve generated by one-factor models are not rich enough to accompany the observations in practice (e.g. Hull-White model with constant or time deterministic mean). The comparison of the yield curve shapes among one-factor models and two-factors stochastic volatility models can be found in Chapter 5. Finally, one-factor models offer poor descriptions of the historical short rate processes, as we will show in Chapter 4.

A number of multifactor models have been presented to overcome the drawbacks of one-factor models. Surveys of the literature about the multifactor models can be found in Strickland (1996) and Rogers (1995), for example. How many factors do

we need, considering both model parsimony and efficiency in explaining the historical observations and pricing the products? Litterman & Scheinkman (1991) found that three common factors (*level*, *steepness* and *curvature*) have the most influence on the returns of the fixed-income products. Based on their studies the first two common factors almost explain 95% or even more variations in return. Therefore, two-factor models would be a good starting point for the extension of short rate models. For the short-rate model with two factors, the empirical work of Dybvig (1995) shows that the other factor is most likely to be the volatility of the short rate and the volatility has a significant effect on the bond option pricing.

This chapter is organized as follows: firstly, we present the Lévy process which is used to compose the volatility part in the short-rate model; secondly we introduce the stochastic volatility models driven by different Lévy processes. We start from the Gaussian driven model; then we present the compound Poisson jump model and the superposition jump model.

2.2 Lévy Process

A *Lévy process* is a cadlag [1] stochastic process $\{Z_t\}_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with $Z_0 = 0$ almost surely, and satisfying the following properties:

1. the increments $\{\Delta Z_{t_i} = Z_{t_i} - Z_{t_{i-1}}\}$, $i = 1, 2, ..., n$, for all the non-overlapping time interval $\Delta t_i = t_i - t_{i-1}$ with $t_0 = 0$, are independent;

2. the distribution of the increments ΔZ_{t_i} are stationary and depend only on the length of Δt_i;

3. $\lim_{\Delta t_i \to 0} P(|Z_{t_i} - Z_{t_{i-1}}|) \geq \epsilon$, $\forall \epsilon$.

Strict definition can be found in Sato (1999).

[1] right continuous with left limits

By the definition above, a Lévy process Z_t is a sum of n i.i.d. random variables. Let f be the density function for ΔZ_{t_i}, then the marginal distribution of Z_t will be the n-th convolution of f and it is *infinitely divisible*. Conversely, for any infinite divisible distribution F, there exists a Lévy process Z_t such that the distribution of Z_1 is given by F (see the Theorem 7.10 of Sato (1999)). This property can help us to build a Lévy process given a particular distribution. Here are some well-known infinite divisible distributions: Gaussian, Poisson, Gamma, inverse Gaussian, and inverse Gamma.

Brownian motion,$\{W_t\}_{t\geq 0}$, is one of the well-known Lévy process with increment, ΔW, following the Normal distribution $N(0, \Delta t)$. Its characteristic function is:

$$\Psi(u) = E(e^{iuW_t}) = \exp\left(-\frac{\Delta t u^2}{2}\right).$$

The marginal law of W_t is also the Gaussian with zero mean and variance t. It is the only Lévy process which has a continuous sample path. Therefore, the model based on a Brownian motion excludes the possibility of jumps in the underlying process. The left plot of Figure 2.2 presents a sample path of a Brownian motion in a unit time $[0, 1]$.

Another fundamental Lévy process is the *compound Poisson process*. A *compound Poisson process* is a continuous-time stochastic process Z_t defined as

$$Z_t = \sum_{i=1}^{N(t)} X_i,$$

where the jumps sizes $\{X_i\}$ are i.i.d. random variables; $N(t)$ is the number of jumps in $(0, t)$ following a Poisson distribution with positive intensity λ, and it is independent from X_i. The characteristic function of Z_t is :

$$\Psi(u) = \exp\left(t \int_{\mathbb{R}} (e^{iuz} - 1)\nu(dz)\right),$$

where ν is the *Lévy measure* [2] of Z_t. For the compound Poisson process, $\nu = \lambda p$ where p is the density function of the jump size X_t. The mean and variance of Z_t

[2]Lévy measure, $\nu(A)$, is the expected number of jumps whose size belongs to a measurable set A in a unit time

are:

$$\begin{aligned} E(Z_t) &= \lambda t E(X_t), \\ Var(Z_t) &= \lambda t E(X_t^2). \end{aligned}$$

Here are some properties of the compound Poisson process:

- it is the only Lévy process with piecewise constant sample paths;
- it is the only Lévy process with almost surely a finite number of jumps in interval $[0, t]$.

The right plot of Figure 2.2 presents the evolution of a compound Poisson process with the exponential distributed jumps.

Generally, if Z_t is a Lévy process on $\mathbb{R}^d$, it can be represented as a superposition of a Brownian motion with drift and a possibly infinite sum of the independent compound Poisson processes:

$$Z_t = \gamma t + AW_t + J_t,$$

where γt represents the drift part with parameter vector γ; W_t is a multidimensional standard Brownian motion and A is a positive definite matrix; and J_t is a jump process with Lévy measure ν. (γ, A, ν) is also called as the *Lévy triplet*. Obviously, a standard Brownian motion is a Lévy process with triplet $(0, 1, 0)$ and the compound Poisson process has the triplet $(0, 0, \nu)$. By the *Lévy-Khinchin representation theorem*, characteristic function of Z_t has the following form:

$$\Psi(u) = \exp\left(i\gamma ut - \frac{1}{2}uAut + t\int (e^{iux} - 1 - iuxI(|x| < 1))\nu(dx)\right),$$

where $I()$ is an indicator function equal to 1 when $|x| < 1$ and zero otherwise.

2.3 Stochastic Volatility Models

The general structure of the two-factor models which we will consider have the form:

$$\begin{aligned} dr_t &= \mu_r(r_t)dt + \sigma_r(r_t, h_t)dW_t, \\ dh_t &= \mu_h(h_t)dt + \sigma_h(r_t, h_t)dZ_t, \end{aligned}$$

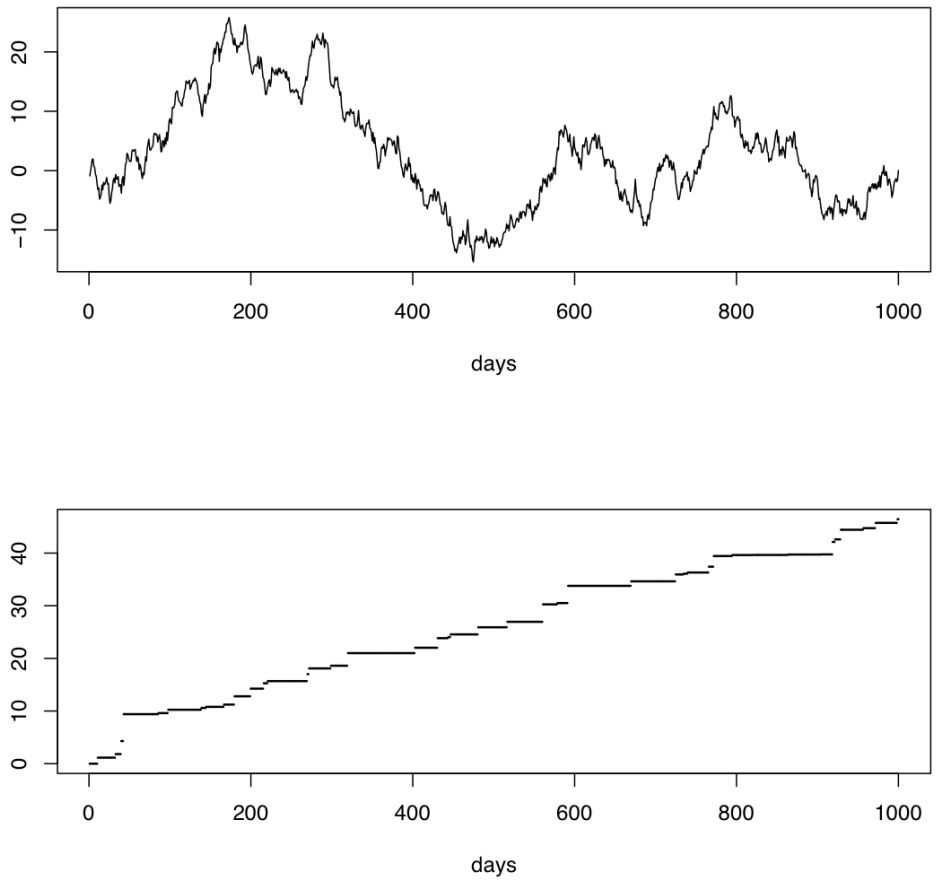

Figure 2.2: Top: The sample path of a Brownian motion in 1000 days. Bottom: The sample path of a compound Poisson process (which intensity rate is 0.05 per day) with the exponential distributed jump size (which intensity rate is 1).

where $\{W_t\}$ and $\{Z_t\}$ are two *Lévy processes* called as the *background driving Lévy process* (BDLP), (Barndorff-Neilsen & Shephard (2001)); r_t is the short rate and h_t is the driving random factor of the *instantaneous volatility process*, $\sigma_r(r_t, h_t)$, of r_t. They are both set as the time homogeneous processes.

For the drift part of the short rate process, we preserve the mean-reversion property by setting $\mu_r(r_t) = \alpha - \beta r_t$. This feature is observed from the historical data of all kinds of interest rate processes. The volatility part is defined as below:

$$\sigma_r(r_t, h_t) = f(h_t) r_t^{\gamma},$$

it is composed of both r_t^{γ} in order to capture the *level effect* of short rate and another random factor to incorporate the heteroskedasticity. Various Lévy processes have been considered as the BDLP for r_t in many literatures, Ahn & Thompson (1988) extend the CIR model by using the Brownian motion accompanied with a Poisson process with constant jump size, Das & Foresi (1996) study the jump size following exponential Bernoulli distribution in Vasicek model and more other distributions for jump size have been tested in Attari (1999). Since we will focus on the effect of the stochastic volatility, we only consider the Brownian motion as the BDLP for r_t. Therefore the short rate model we will consider has the form:

$$dr_t = (\alpha - \beta r_t)dt + f(h_t) r_t^{\gamma} dW_t.$$

This model is then a CKLS model with a stochastic volatility.

The nonnegative volatility process $\sigma_r(r_t, h_t)$ can be obtained by selecting the proper function form $f()$ or a positive state process h_t. Under the general structure, many models have been considered to build h_t: the geometric Brownian motion used in Hull & White (1987); the CIR model applied by Heston (1993) and Ornstein-Uhlenbeck (OU) process Andersen & Lund (1997). Here we will consider h_t as the OU process driven by a continuous or jump Lévy processes. Generally the h_t process is given by:

$$dh_t = (\kappa - \mu h_t)dt + \sigma_h dZ_t, \tag{2.23}$$

if Z_t is a Brownian motion, then h_t follows the *Gaussian OU* process, and if Z_t is any other Lévy process, h_t follows the *non-Gaussian OU* process. The model is therefore designed to capture the observed features of short rate, the *fat tail*, the *skewness* and the *volatility clustering* for example. In following sections, we will discuss the different forms of the positive function $f()$ and the various driving processes in h_t.

2.4 SV Model Driven by Wiener Process: SVG

First, we consider the SVG model where h_t is a Gaussian OU process. Inspired by the success of similar asset return models (Scott (1997)), we suppose that the volatility is an exponential function of h_t given $\{r_t\}$

$$f(h_t) = \sigma_r \exp(h_t/2),$$

which is a positive function with $\sigma_r > 0$. h_t is a mean-reverting OU process driven by a Brownian motion, Z_t:

$$dh_t = (\kappa - \mu h_t)dt + \sigma_h dZ_t,$$

where κ, μ and σ_h are positive constant parameters; μ is the mean-reverting rate, κ/μ is the long-run average level of h_t and Z_t is independent with W_t, the Brownian motion in r_t. Andersen & Lund (1997) compare this model with the GARCH and the EGARCH models. They provide the consistent estimation for the parameters from the Efficient Method of Moment (EMM). They find that there are enormous improvements in data fitting with the introduction of the stochastic volatility factor. However, they also point out that the failure of the SVG model in generating innovations with sufficiently fat tails.

Let $\hat{h}_t = h_t - \frac{\kappa}{\mu}$ and $\hat{\sigma}_r = \sigma_r \exp(\frac{\kappa}{2\mu})$, then the SVG model which we will consider has the following specification:

$$dr_t = (\alpha - \beta r_t)dt + \hat{\sigma}_r \exp(\hat{h}_t/2) r_t^{\gamma} dW_t, \tag{2.24}$$

$$d\hat{h}_t = -\mu \hat{h}_t dt + \sigma_h dZ_t. \tag{2.25}$$

This adjusted structure will improve the convergence rate of MCMC (see Eraker (2001) for details). Figure 2.3 presents the simulated sample paths of h_t, $\exp(h_t)$ and the corresponding short rate r_t.

We can apply the *infinitesimal generator* method (Cont & Tankov (2004)) to deduce the stationary distribution of σ_t. From Equation (2.25) we can deduce

$$\hat{h}_t = \hat{h}_0 e^{-\mu t} + \sigma_h \int_0^t e^{-\mu(t-s)} dZ_s.$$

Set $f(\hat{h}) = e^{iu\hat{h}}$, the *infinitesimal generator* of $\hat{h}_t$ for function $f()$ is given by:

$$\begin{aligned} Lf(\hat{h}) &= \lim_{\Delta t \to 0} \frac{E(f(\hat{h}_{t+\Delta t}) - f(\hat{h}_t))}{\Delta t} \\ &= \frac{\sigma_h^2}{2} \frac{\partial^2 f(\hat{h})}{\partial \hat{h}^2} - \mu \hat{h} \frac{\partial f(\hat{h})}{\partial \hat{h}} \\ &= -\frac{1}{2}\sigma_h^2 u^2 f(\hat{h}) - \mu \hat{h} i u f(\hat{h}). \end{aligned}$$

If F_t is the distribution of $\hat{h}_t$, by the Kolmogorov backward equation (Fokker-Planck equation) we can get:

$$\frac{d}{dt} E_t(f) = E_t(Lf),$$

where $E_t(f) = \int f(\hat{h}) dF_t$. If F is the stationary distribution, the equation below will be obtained:

$$\begin{aligned} E_t(Lf) &= -\int \frac{1}{2}\sigma_h^2 u^2 f(\hat{h}) + \mu \hat{h} i u f(\hat{h}) dF_t \\ &= -\frac{1}{2}\sigma_h^2 u^2 \Psi(u) - \mu u \Psi(u)' \\ &= 0, \end{aligned}$$

where $\Psi(u) = E(e^{iu\hat{h}})$ is the characteristic function. Then $\Psi()$ can be calculated from above equation and has the form:

$$\Psi(u) = \exp\left\{-\frac{u^2\sigma_h^2}{4\mu}\right\}$$

which is the characteristic function of Normal distribution with the variance equals to $\frac{\sigma_h^2}{2\mu}$ and zero mean. Then $\hat{h}_t \sim N(0, \frac{\sigma_h^2}{2\mu})$ as $t \to \infty$. Due to the exponential relation, the stationary probability law of short rate volatility σ_t given r_t is a log-normal

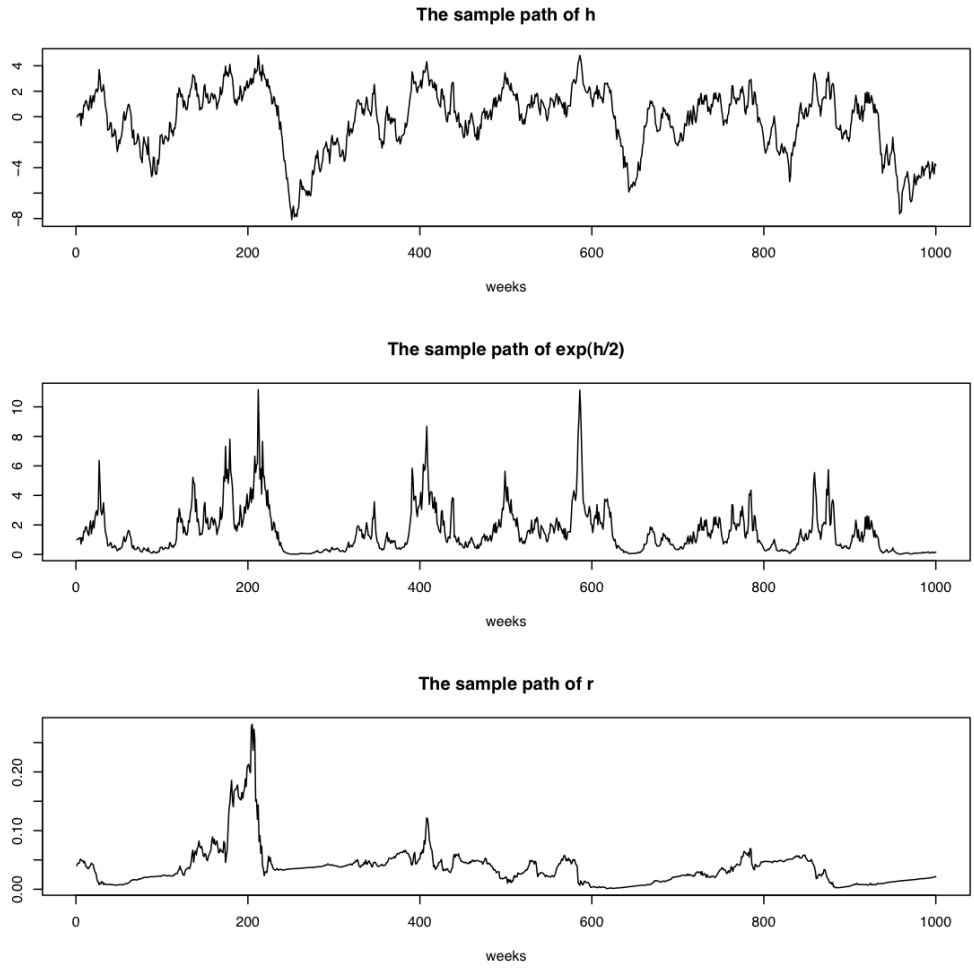

Figure 2.3: Top: The sample path of the volatility driven process h_t in 1000 weeks. Middle: The sample path of the exponential process of h_t in 1000 weeks. Bottom: The sample path of the corresponding short rate process in 1000 weeks. The simulations are generated from Equation (2.24) and (2.25). The parameter values used are: $\alpha = 0.0002, \beta = 0.003$, $\gamma = 0.8$, $\sigma_r = 0.03$, $\mu = 0.07$, $\sigma_h = 0.6$, $h_0 = 0$ and $r_0 = 0.04$.

distributions with mean $\sigma_r r_t^{\gamma} e^{\frac{\sigma_h^2}{4\mu}}$ and variance $\sigma_r^2 r_t^{2\gamma} e^{\frac{\sigma_h^2}{2\mu}}(e^{\frac{\sigma_h^2}{2\mu}} - 1)$.

The MCMC procedure for parameter estimations is given in Section 3.3.2. The estimating results are listed in Table 4.9. Due to the complexity of the model structure, we cannot get any analytical pricing formulae. We will apply Monte Carlo simulation which will be presented in Chapter 5.

2.5 SV Model Driven by Compound Poisson Process

As we will point out in Chapter 4, the conditional density of the residuals from the SVG model still have fatter tails than the Normal distribution. One way to improve the model is to introduce further factors, stochastic drift for example. Andersen & Lund (1997) find that neither stochastic drift nor stochastic volatility alone work well until they are combined. However, they also mentioned the problem of accommodating extreme points of the short rate innovations with this three-factor model. Instead of introducing more factors to explain the dynamics of short rate, we will investigate more general stochastic driven processes in the volatility part.

Since the Brownian motion is just one special case of the Lévy process, other Lévy processes are worth examining to compensate the drawback of Gaussian driven process. We turn to the stochastic volatility models driven entirely by positive jump process.

Let the random factor of volatility be an OU process but driven by a jump process:

$$dh_t = -\mu h_t dt + \sigma_h dZ_t,$$

where Z_t is a compound Poisson process with intensity rate λ. Z_t is independent of

W_t. It is defined by:

$$Z_t = \sum_{j=1}^{N(t)} \varepsilon_j I(c_j \leq t),$$

where $0 < c_1 < ... < c_{N(t)} \leq t$ are the jump times in $(0, t)$; $I(.)$ is indicator function; $N(t)$ is the total number of jumps up to and including time t and the $\{\varepsilon_j\}$ are *i.i.d.* variables as the jump size following a standard exponential distribution, which are also independent of $\{c_j\}$. h_t is therefore a positive OU process if $h_0 > 0$. The quadratic function form can be adopted for the short rate volatility:

$$f(h_t) = h_t^{1/2}.$$

This model was first introduced by Barndorff-Neilsen & Shephard (2001) for the volatility of asset returns modeling. Then the SVJ model has the following specification:

$$dr_t = (\alpha - \beta r_t)dt + h_t^{1/2} r_t^{\gamma} dW_t, \tag{2.26}$$

$$dh_t = -\mu h_t dt + \sigma_h dZ_t. \tag{2.27}$$

Figure 2.4 presents the simulated sample paths of the jump points $(c_i, \varepsilon_i)_{i=1}^{N(T)}$, h_t and the corresponding short rate process r_t.

One of the advantages of such a jump volatility model is analytic tractability of the integrated volatility which is particular useful in both theory and practice. As we will see, we apply the integrated volatility in the MCMC estimating procedure. Define the integrated volatility as:

$$v(s, t) = \int_s^t h_u du. \tag{2.28}$$

Integrate both sides of Equation (2.27) directly from time s to t, we get:

$$h_t - h_s = -\mu \int_s^t h_u du + \sigma_h (Z_t - Z_s), \tag{2.29}$$

and with Equation (2.28), we can obtain the the following specification for the SVJ model :

$$v(s, t) = \frac{1}{\mu}(\sigma_h \sum_{j=1}^{N(t-s)} \varepsilon_{j+N(s)} - (h_t - h_s)), \tag{2.30}$$

where $N(t-s)$ is the number of jumps in (s,t) and (c_j, ε_j) is the jump point. The specification of h_t in above equation can also be obtained from Equation (2.27) by multiplying $e^{-\mu t}$ on both sides before integration:

$$\begin{aligned} h_t &= h_0 e^{-\mu t} + \sigma_h \int_0^t e^{-\mu(t-s)} dZ_s \\ &= h_0 e^{-\mu t} + \sigma_h \sum_{j=1}^{N(t)} e^{-\mu(t-c_j)} \varepsilon_j, \end{aligned} \tag{2.31}$$

where $N(t)$ is the number of jumps in $(0,t)$ and (c_j, ε_j) is the jump point.

Stationary distribution

From Equation (2.31), the value of h_t increases by jump but decreases exponentially with decay rate μ. We will find The characteristic function of h_t by Lemma 2.1.

Lemma 2.1 *Let $f : [0,T] \to R$ be left-continuous and $(Z_t)_{t\geq 0}$ be a Lévy process. Then*

$$E\left\{\exp\left(i\int_0^T f(t)dZ_t\right)\right\} = \exp\left\{\int_0^T \psi(f(t))dt\right\}$$

where $\psi(u)$ is the characteristic exponent of Z.

Proof can be found in Cont & Tankov (2004). Given Equation (2.31), the characteristic function of h_t is as below:

$$\begin{aligned} E\left(e^{iuh_t}|h_0\right) &= E\left(\exp\left\{iuh_0e^{-\mu t} + iu\sigma_h \int_0^t e^{-\mu(t-s)} dZ_s\right\}\right) \\ &= \exp\left\{iuh_0e^{-\mu t} + \int_0^t \psi\left(\sigma_h u e^{-\mu(t-s)}\right) ds\right\}, \end{aligned}$$

where $\psi()$ is the characteristic exponent of Z_t. Since the characteristic exponent for a compound Poisson with exponential jump size is $\psi(x) = \frac{i\lambda x}{1-ix}$. Then characteristic function of h_t has the following form:

$$\begin{aligned} E\left(e^{iuh_t}|h_0\right) &= \exp\left\{iuh_0e^{-\mu t} + \int_0^t \frac{iue^{-\mu(t-s)}}{\sigma_h^{-1} - iue^{-\mu(t-s)}} ds\right\} \\ &= \exp\left\{iuh_0e^{-\mu t} - \frac{\lambda}{\mu}\log\left(\frac{\sigma_h^{-1} - iu}{\sigma_h^{-1} - iue^{-\mu t}}\right)\right\} \\ &= \exp\left\{iuh_0e^{-\mu t}\right\}\left(\frac{1-\sigma_h iue^{-\mu t}}{1-iu\sigma_h}\right)^{\nu}. \end{aligned} \tag{2.32}$$

The stationary characteristic function is:

$$\begin{aligned} E\left(e^{iuh}\right) &= \lim_{t\to+\infty} E\left(e^{iuh_t}|h_0\right) \\ &= \left(\frac{1}{1-iu\sigma_h}\right)^{\nu} \end{aligned}$$

where $\nu = \lambda/\mu$. Therefore the stationary distribution of h is a gamma distribution, $\Gamma(\nu, \sigma_h^{-1})$, with mean $\nu\sigma_h$ and variance $\nu\sigma_h^2$.

More generally, there is a close link between the OU processes and a class of infinitely divisible distributions called *self-decomposable distributions*. A random variable X is *self-decomposable* if for every $b > 1$ there exists a random variable ϵ_b independent of X such that

$$\Psi_X(u) = \Psi_X(u/b)\Psi_{\epsilon_b}(u),$$

where $\Psi_X(u)$ is the characteristic function of X and $\Psi_{\epsilon_b}(u)$ is the characteristic function of ϵ_b. The following theorem, which is the proposition 15.4 of Cont & Tankov (2004), describes the relationship of OU process and corresponding stationary distribution.

Theorem 2.1 *Let $\{Z_t\}$ be a Lévy process with characteristic triplet (A, ν, γ). If*

$$\int_{|x|\geq 1} \ln|x|\nu(dx) < \infty$$

then the OU process $\{y_t\}$ driven by $\{Z_t\}$ has a stationary distribution f which is self-decomposable. Conversely, for every self-decomposable distribution f there exists a Lévy process $\{Z_t\}$ such that f is the stationary distribution of the OU process driven by $\{Z_t\}$.

Proof can be found in Cont & Tankov (2004). This theorem works as a guideline for us to construct an OU process through the stationary distribution. Both Gaussian and gamma distribution are self-decomposable, and more examples have been considered in Barndorff-Neilsen & Shephard (2001).

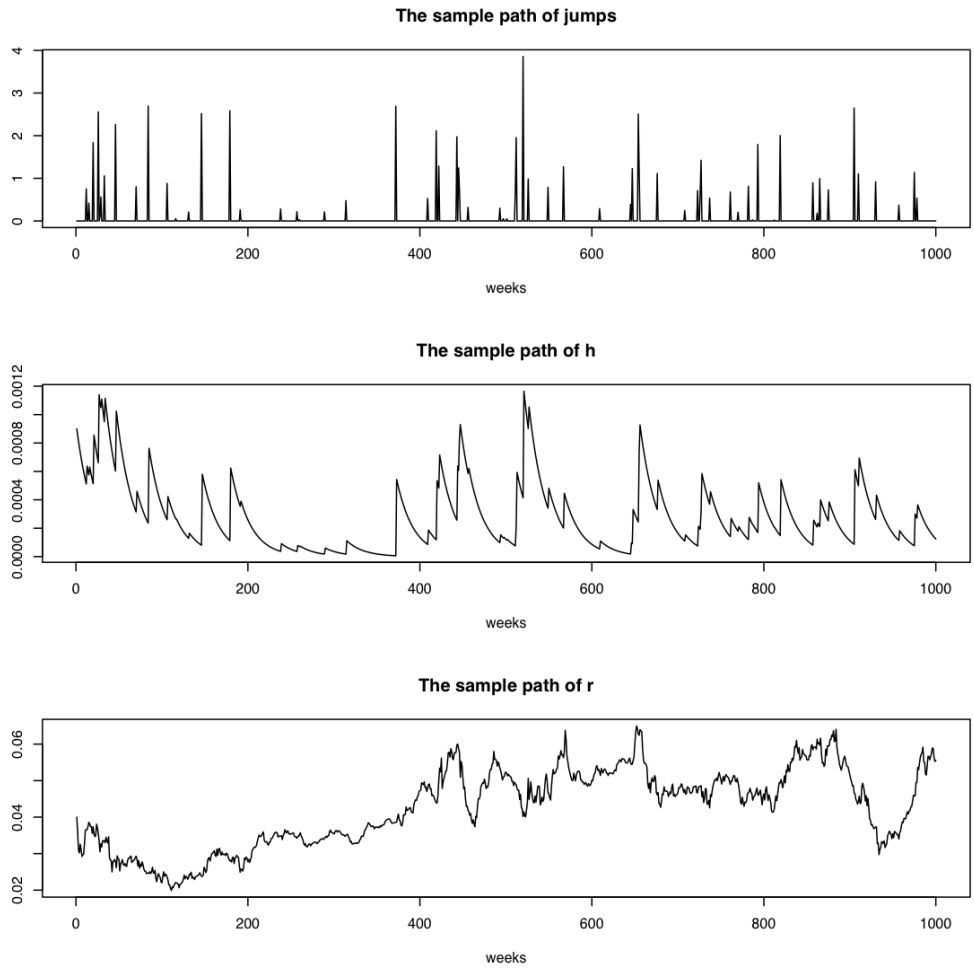

Figure 2.4: Top: The sample path of the jumps in 1000 weeks. Middle: The sample path of the volatility driven process h_t in 1000 weeks. Bottom: The sample path of the corresponding short rate process in 1000 weeks. The simulations are generated from Equation (2.26) and (2.27). The parameter values used are: $\alpha = 0.0002, \beta = 0.003$, $\gamma = 0.8$, $\mu = 0.05$, $\nu = 1.5$, $\sigma_h = 0.0002$, $h_0 = 0.0009$ and $r_0 = 0.04$.

Moment functions

The moment function of h_t can be found from its characteristic function if $E(|h_t|^n) < \infty$:

$$E(h_t^k) = \frac{1}{i^k}\frac{\partial^k \Psi}{\partial u^k}(0),$$

for $k = 1, 2, ..., n$. Given Equation (2.32), the mean and variance of h_t are

$$\begin{aligned} E(h_t|h_0) &= h_0 e^{-\mu t} - \nu\sigma_h e^{-\mu t} + \nu\sigma_h, \\ Var(h_t|h_0) &= \nu\sigma_h^2(1 - e^{-2\mu t}). \end{aligned}$$

In the stationary case, the autocorrelation of h_t is

$$\begin{aligned} acf(s) &= \frac{cov(h_t, h_{t+s})}{\sqrt{Var(h_t)Var(h_{t+s})}} \\ &= \frac{E(h_t h_{t+s}) - E(h_t)^2}{Var(h_t)} \\ &= \frac{E(h_t E(h_{t+s}|h_t)) - E(h_t)^2}{Var(h_t)} \\ &= \frac{E(h_t(h_t e^{-\mu s} - \nu\sigma_h e^{-\mu s} + \nu\sigma_h)) - \nu^2\sigma_h^2}{\mu\sigma_h^2} \\ &= e^{-\mu s}, \end{aligned} \tag{2.33}$$

which is only related to the decay parameter μ. For all the stationary OU process defined as Equation (2.23), the autocorrelation is always $e^{-\mu s}$. More complicated dependence structures can be obtained by adding together independent OU processes as suggested by Barndorff-Neilsen & Shephard (2001) and in the next section, we will consider the superposition of non-Gaussian OU processes with Gamma marginal distribution.

2.6 SV model with superposition of jumps: SVS

As shown in Equation (2.33), the dependence structure implied from the OU model is an exponential decaying process. Whereas, it is still not consistent with the structures observed from the market, Barndorff-Neilsen (2001). The typically empirical

dependence structure is relatively fast decline in previous lags, followed by a slower decay for the rest duration.

The SVJ model can be improved to incorporate this problem by substituting the single OU process for a sum of independent OU processes, see Barndorff-Neilsen (2001) and Barndorff-Neilsen & Shephard (2001). The random factor h_t, of volatility will be:

$$h_t = \sum_{i=1}^{m} h_{i,t},$$

where $h_{i,t}, i = 1, 2, ..., m$ are independent OU processes with the SDE as follows:

$$dh_{i,t} = -\mu_i h_{i,t} dt + dZ_{i,t},$$

and $Z_{i,t}$ are Lévy processes. Parameter μ_i plays an important role in this model to capture the different decaying rates of the correlations of short rate in different periods. The autocorrelation function of $h_.$ is then:

$$\begin{aligned} acf(s) &= \frac{\sum_{i=1}^{m} e^{-\mu_i s} Var(h_{i,t})}{\sum_{i=1}^{m} Var(h_{i,t})} \\ &= \sum_{i=1}^{m} w_i e^{-\mu_i s}, \end{aligned}$$

setting the weights $w_i = Var(h_{i,t}) / \sum_{j=1}^{m} Var(h_{j,t})$.

Empirical findings in Barndorff-Neilsen & Shephard (2002) suggest that $m = 2$ is adequate for modeling the real equity market data. Then we can capture both short-term variation by the OU process with higher decay rate μ_1, and long-term variation by the one with smaller μ_2. There are much more flexibility in choosing the stationary distributions for $h_{i,t}$ which can belong to same family or not. To compare the capability of models, the SV model with superposition is construct by a sum of two independent gamma-OU processes. The BDLP, $Z_{i,t}$, for each process will be:

$$Z_{i,t} = \sum_{t=0}^{\infty} \varepsilon_{i,j} I(c_{i,j} < t),$$

where $Z_{i,t}$ are independent Compound Poisson processes with different jump time rates λ_i but the jump size $\varepsilon_{i,j}$ is from $Exp(\sigma_h^{-1})$ for both processes. Since the stationary distribution for each $h_{i,t}$ is still a Gamma distribution with parameters $\nu_i = \lambda_i/\mu_i$ and σ_h^{-1} as we have proven in section 2.5, and the distribution of the sum of independent random variables with Gamma distributions are still Gamma distribution, h_t will have stationary distribution: $\Gamma(\sum_{i=1}^{2} \nu_i, \sigma_h^{-1})$. Figure 2.5 and 2.6 present the simulated sample paths of the jump points $(c_{1,i}, \varepsilon_{1,i})_{i=1}^{N_1(T)}$, the jump points $(c_{2,j}, \varepsilon_{2,j})_{j=1}^{N_2(T)}$, $h_{1,t}$, $h_{2,t}$, h_t and the corresponding short rate process r_t.

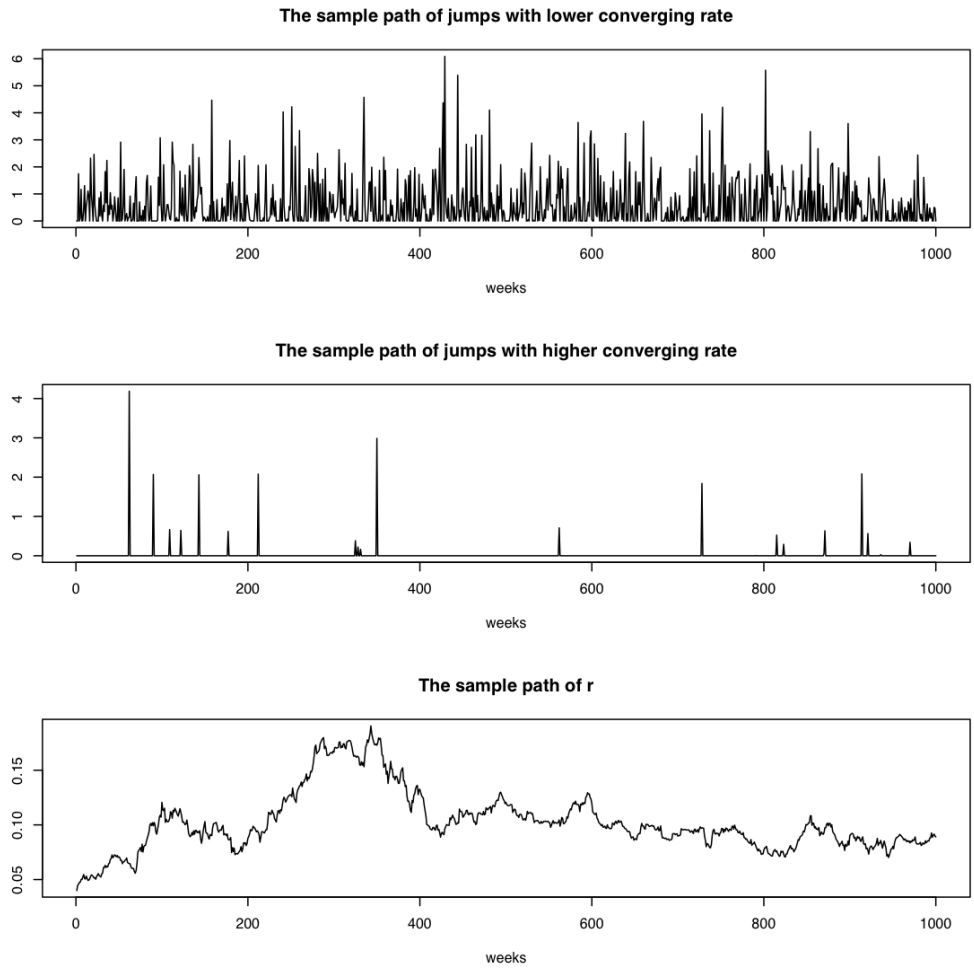

Figure 2.5: Top: The jump path of $h_{1,t}$ in 1000 weeks. Middle: The jump path of $h_{2,t}$ in 1000 weeks. Bottom: The sample path of the corresponding short rate process in 1000 weeks. The simulations are generated from Equation (2.26) and (2.27). The parameter values used are: $\alpha = 0.0002, \beta = 0.003$, $\gamma = 0.8$, $\mu_1 = 0.7$, $\nu_1 = 1.5$, $\mu_2 = 0.03$, $\nu_2 = 1$, $\sigma_h = 0.0002$, $h_{1,0} = 0.00045$, $h_{2,0} = 0.00045$ and $r_0 = 0.04$.

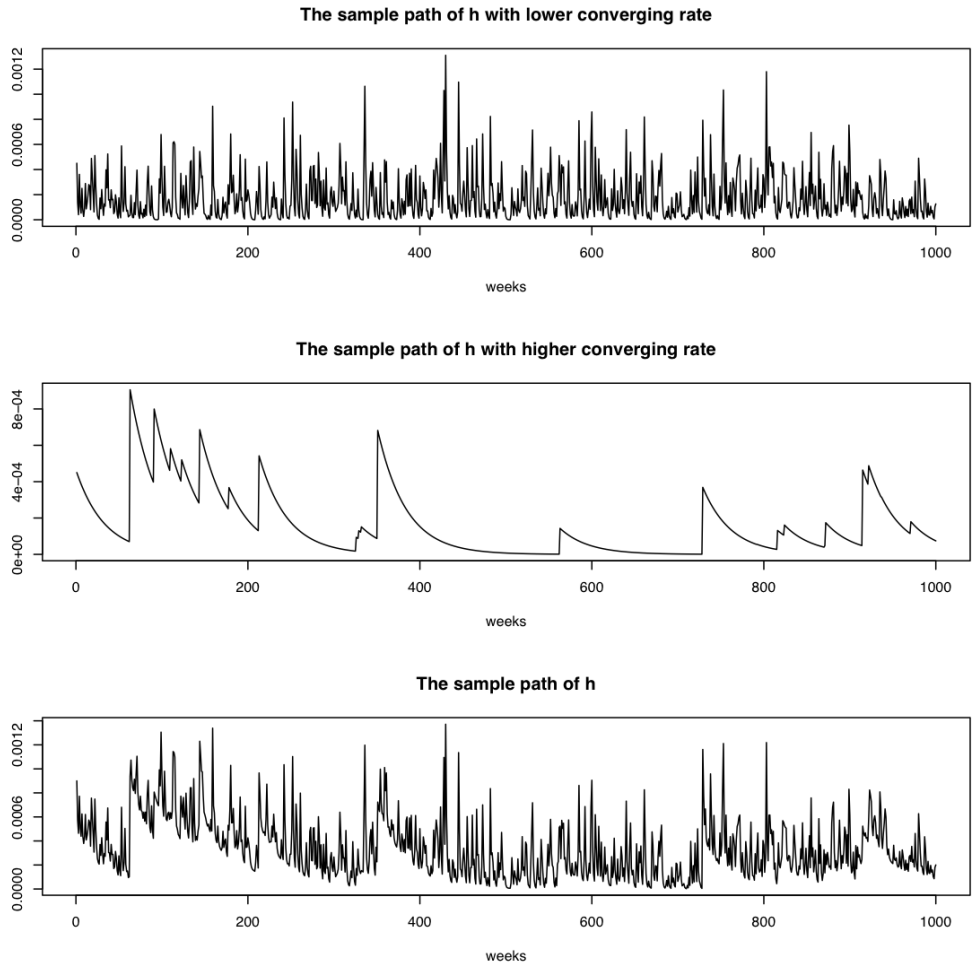

Figure 2.6: Top: The sample path of the volatility driven process $h_{1,t}$ in 1000 weeks. Middle: The sample path of the volatility driven process $h_{2,t}$ in 1000 weeks. Bottom: The sample path of the volatility driven process h_t in 1000 weeks. The simulations are generated from Equation (2.26) and (2.27).

Chapter 3

MCMC Algorithm for Parameter Estimations

3.1 Introduction

We have presented both the one-factor CKLS model and the stochastic volatility models in previous chapters. In this chapter we will develop the Markov chain Monte Carlo algorithms to estimate the parameters of these candidate models. We begin our description from the simplest structure, the CKLS model. The MCMC procedures for the SV models are presented in following sections. Thesimulation tests are provided in latter sections for each model to assess the efficiency of the corresponding algorithms.

3.2 MCMC for the CKLS Model

3.2.1 Introduction

To estimate the parameters of the continuous model given the discontinuous market data, we should find the discrete version of the model. For the CKLS model, Chan *et al.* (1992) apply the *Euler approximation* which discretizes the model as below:

$$\Delta r_i = (\alpha + \beta r_i)\Delta t + \sigma_r r_i^{\gamma} \Delta W_i, \tag{3.34}$$

where r_i is the short rate at the ith time step, $\Delta t = t_{i+1} - t_i$ (one day, week, or month, for example); $\Delta r_i = r_{i+1} - r_i$ and ΔW_i is a normal variable with the variance, t. Chan *et al.* (1992) work with the monthly data where $\Delta t = 1$ (month). Eraker (2001) (who applies MCMC) uses the weekly data first and then set Δt smaller than the observation interval. The missing data between two neighbour observations are treated as unknown parameters which will be simulated from their posterior distributions during the MCMC iteration. This is known as the *data augmentation algorithm*, details are in Section 3.3.1.

For the CKLS model, the unknown parameter vector is $\Theta = (\alpha, \beta, \sigma_r, \gamma)'$. To estimate these variables with MCMC algorithm, we first need to figure out the structures of the posterior distributions.

3.2.2 Posterior distribution of the CKLS model

According to the discrete specification (3.34), the conditional joint density of $\{r_i\}$ given $\Theta = (\alpha, \beta, \sigma_r, \gamma)'$, the likelihood function $L(r|\Theta)$ is proportion to

$$L(r|\Theta) \propto \prod_{i=0}^{n-1} \sigma_r^{-1} \Delta t^{-0.5} r_i^{-\gamma} \exp\left\{ -\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\sigma_r^2 r_i^{2\gamma} \Delta t} \right\}. \tag{3.35}$$

From the Bayes theorem, the joint posterior density for unknown parameters are proportion to the product of the joint prior density and the likelihood function

$$p(\Theta|r) \propto L(r|\Theta) p(\Theta), \tag{3.36}$$

where $p(\Theta)$ is the joint prior distribution of the parameters.

The dependence of the observation and the parameters can be interpreted more clearly by the *directed acyclic graph* (DAG). In this graph, the rectangle shape of node represents an observation, and the circle one represents an unknown factor. The nodes are linked by arrows. The one where an arrow emanating is defined as the *parent* of the node which the arrow pointing to, and the pointed node is the *descendant*. Two types of arrows are used here. The one with solid line indicates

the deterministic dependence between the nodes and the probabilistic relation is presented by the dashed line. In such graph model, if we know the value of one node's parents, then no others would be informative to it except its descendants. Then the joint distribution of all the random variables is fully determined by the conditional distribution of each node given its parent. More definitions and properties of the DAG can be found in Gilks *et al.* (1996). The graphic method is very helpful to identify the structure of the model, especially for a complex model. The DAG of the CKLS model is shown in Figure 3.7.

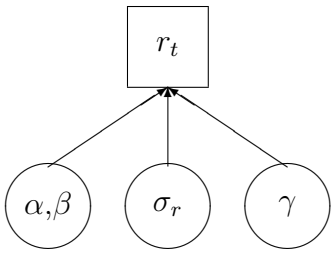

Figure 3.7: Graphical model for the CKLS model

With little information about the parameters of the CKLS model, we use uniform distributions within different interval as the priors to these parameters. The effect of the prior distribution on the acceptance probability would be smaller by selecting wider range of value interval. Then we assume $\alpha \in (0, 0.1)$, $\beta \in (-0.1, 0)$, $\gamma \in (0, 2)$ and $\sigma_r \in (0, 0.5)$. Considering Equation (3.35) , (3.36) and the DAG of the CKLS model, the posteriors for each parameter will be

$$p((\alpha,\beta)|r,\Theta_{-(\alpha,\beta)}) \propto \prod_{i=0}^{n-1} \exp\left\{-\frac{(\Delta r_i-(\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 r_i^{2\gamma}\Delta t}\right\}, \tag{3.37}$$

$$p(\sigma_r|\Theta_{-\sigma_r}) \propto \prod_{i=0}^{n-1} \sigma_r^{-1} \exp\left\{-\frac{(\Delta r_i-(\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 r_i^{2\gamma}\Delta t}\right\}, \tag{3.38}$$

$$p(\gamma|\Theta_{-\gamma}) \propto \prod_{i=0}^{n-1} r_i^{-\gamma} \exp\left\{-\frac{(\Delta r_i-(\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 r_i^{2\gamma}\Delta t}\right\}, \tag{3.39}$$

where, for example, $\Theta_{-(\alpha,\beta)'} = (\sigma_r, \gamma)'$ is the parameter vector comprising all members of Θ except for $(\alpha, \beta)'$.

3.2.3 Sample algorithm for parameters of the CKLS model

(α, β)

We combine both α and β in one section since we can deduce an explicit joint posterior function for them. Setting $y = (y_1, ..., y_n)'$ where $y_i = \Delta r_i r_i^{-\gamma} \Delta t$, $x_i = (r_i^{-\gamma}\sqrt{\Delta t}, r_i^{1-\gamma}\sqrt{\Delta t})'$ and $X = (x_1, ...x_i, ..., x_n)'$ is a $n \times 2$ matrix, the joint posterior of (α, β) from Equation (3.37) can be rewritten as:

$$p((\alpha,\beta)|r, \Theta_{-(\alpha,\beta)}) \propto \exp\left\{-\frac{(y'y - 2(\alpha,\beta)X'y + (\alpha,\beta)X'X(\alpha,\beta)')}{2\sigma_r^2}\right\}, \quad (3.40)$$

which is a Normal distribution with mean, $M_1 = (X'X)^{-1}(X'y)$ and variance $S_1^2 = \sigma_r^2(X'X)^{-1}$. Then we can sample (α, β) directly from the Normal posterior distribution.

σ_r

The posterior for σ_r^2 is also a well-known distribution. From Equation (3.38), we can deduce:

$$p(\sigma_r|\Theta_{-\sigma_r}) \quad \propto \quad (\sigma_r^{-2})^{n/2} \exp\left\{\sigma_r^{-2} \sum_{i=0}^{n-1} -\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2r_i^{2\gamma}\Delta t}\right\} \quad (3.41)$$

which indicates that σ_r^2 follows an inverse Gamma distribution, $IG(n/2-1, A)$ with $A = \sum_{i=0}^{n-1} \frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2r_i^{2\gamma}\Delta t}$. We can obtain the sample of σ_r by simulating σ_r^{-2} from Gamma distribution, $G(n/2-1, A^{-1})$.

γ

Whereas, the posterior distribution of γ does not belong to any well-known family, the Metropolis-Hastings algorithm can be used to generate the sample. We use a Normal distribution as the proposal distribution to simulate the candidate value ξ_γ. The converging rate can be adjusted by changing the value of the variance of the proposal distribution, S_2^2. Setting the previous state value of γ as the mean, this methodology is the so-called *random walk* MH algorithm. Accompanied by

Equation (3.39), the acceptance probability for γ is:

$$\begin{aligned} H(\gamma, \xi_\gamma) &= \min\left(1, \frac{P(\xi_\gamma|\Theta_{-\gamma})}{P(\gamma|\Theta_{-\gamma})}\right) \\ &= \min\left(1, \prod_{i=0}^{n-1} r_i^{(\gamma-\xi_\gamma)} \exp\left\{-A_i\left(r_i^{-2\xi_\gamma} - r_i^{-2\gamma}\right)\right\}\right) \end{aligned} \tag{3.42}$$

where $A_i = \frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 \Delta t}$. Then $\gamma = \xi_\gamma$ with probability $H(\gamma, \xi_\gamma)$ and $\gamma = \gamma$ with probability $1 - H(\gamma, \xi_\gamma)$.

The MCMC algorithm for the CKLS model sampling is:

step 1 Initialize Θ at beginning ;

step 2 update (α, β) using Gibbs step from the Normal distribution as Equation (3.40);

step 3 update σ_r using Gibbs step from the inverse Gamma posterior distribution as Equation (3.41);

step 4 update γ using M-H step with the acceptance probability given by Equation (3.42);

step 5 go back to step 2.

Before we apply this MCMC method to the CKLS model with historical data, the efficiency test with predetermined value of parameters will be given in the latter Section 3.5. Theempirical results and statistical tests of these parameters are provided in Chapter 4.

3.3 MCMC for the SVG Model

3.3.1 Data Augmentation for the Latent Process

Since the stochastic volatility part is a latent process, methods based solely on the historical observations are not the proper candidate to work with. One way to handle unobservable or missing data is to apply the MCMC algorithm based on the

data augmentation principle, introduced in statistical inference by Tanner & Wang (1996).

Let Y be the observable data set, Θ be the unknown parameter vector and Z be the missing data or latent data set. The general idea for data augmentation can be represented by the following equation:

$$p(\Theta|Y) = \int p(\Theta|Y,Z)p(Z|Y)dZ. \tag{3.43}$$

The posterior of Θ can be obtained from the density of Θ conditioned on both observed and latent data, by integrating out the unobserved data. From equation (3.43), the posterior can also be written as the expectation:

$$p(\Theta|Y) = \int p(\Theta|Y,Z)dP(Z|Y) = E_{P(Z|Y)}\left(p(\Theta|Y,Z)\right),$$

which can be calculated by approximation:

$$p(\Theta|Y) \approx \frac{1}{n}\sum_{i=1}^{n} p(\Theta|Y,Z_i),$$

and $\{Z_i, i = 1,..,n\}$ are the sample simulated from $p(Z|Y)$.

In our case, we combine the parameters and the latent process as $(\Theta, \{h_t\})$, and construct Markov chains to sample the jth parameter values, $\Theta^{(j)}$ given $\{h_{i\Delta t}\}^{(j-1)}, \{r_i\}$, and then $\{h_{i\Delta t}\}^{(j)}$ given $\Theta^{(j)}, \{r_i\}$ iteratively. The estimation of Θ will be obtained by the Monte Carlo integration over those samples.

3.3.2 Posterior distributions for the SVG Model

The discrete specifications for the SVG model is:

$$\Delta r_i = (\alpha + \beta r_i)\Delta t + \sigma_r e^{\frac{1}{2}\hat{h}_i} r_i^{\gamma} \Delta W_i, \tag{3.44}$$

$$\Delta \hat{h}_i = -\mu \hat{h}_i \Delta t + \sigma_{\hat{h}} \Delta Z_i, \tag{3.45}$$

where $\hat{h}_i$ is defined in Section 2.4, ΔW_i and ΔZ_i are two independent Brownian motions with zero mean and variance Δt for all i.

Figure 3.8 shows the relationships of the observed data, latent process and unknown parameters for the SVG model. Here the random factor $\hat{h}_i$ is conditional independent of the parameters of the short rate part, say $\Theta_1 = (\alpha, \beta, \sigma_r, \gamma)$. Let $\Theta_2 = (\mu, \sigma_{\hat{h}})$ be the vector including the parameters of $\hat{h}_t$, then $\Theta = (\Theta_1, \Theta_2)$.

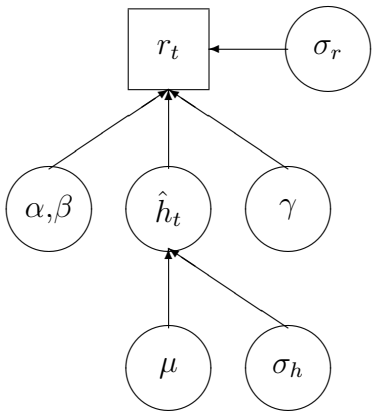

Figure 3.8: Graphical model for the SVG model

The posterior distribution regarding both the parameters and underlying stochastic process will be:

$$p(\Theta, \hat{h}|r) \propto L(r|\Theta_1, \hat{h}) L(\hat{h}|\Theta_2) p(\Theta),$$

where $p(\Theta)$ is the joint prior distribution of Θ. The likelihood function for $\{r_i\}$ can obtained from Equation (3.44) as:

$$L(r|\Theta_1, \hat{h}) \propto \prod_{i=0}^{n-1} \sigma_r^{-1} e^{-\frac{1}{2}\hat{h}_i} r_i^{-\gamma} \Delta t^{-1/2} \exp\left\{ -\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\sigma_r^2 e^{\hat{h}_i} r_i^{2\gamma} \Delta t} \right\}, \tag{3.46}$$

and the likelihood for $\hat{h}$ deduced from Equation (3.45) is:

$$L(\hat{h}|\Theta_2) \propto \prod_{i=0}^{n-1} \sigma_{\hat{h}}^{-1} \Delta t^{-1/2} \exp\left\{ -\frac{\left(\Delta \hat{h}_i + \mu \hat{h}_i \Delta t\right)^2}{2\sigma_{\hat{h}}^2 \Delta t} \right\}. \tag{3.47}$$

Given the process $\{\hat{h}_i\}$, the MCMC algorithm for Θ_1 will be similar to the CKLS model. Instead of the constant volatility parameter σ_r in the CKLS model, we will

update the volatility of the SVG model at each iteration with $\hat{h}_i^{(j)}$ for every time $i\Delta t$. For Θ_2, the Markov chains can be constructed based upon Equation (3.47) as we have the sample of $\{\hat{h}_i\}$.

We use the uninformative uniform prior to each parameter and assume they are independent. The joint posterior density will have the form as:

$$\begin{aligned} p(\Theta|r) &\propto p(\Theta_1|r,\hat{h})p(\Theta_2|\hat{h}) \\ &\propto L(r|\Theta_1,\hat{h})L(\hat{h}|\Theta_2) \\ &\propto \prod_{i=0}^{n-1} \sigma_r^{-1}\sigma_{\hat{h}}^{-1} e^{-\frac{1}{2}\hat{h}_i} r_i^{-\gamma}\Delta t^{-1} \\ &\quad \exp\left\{ -\frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 e^{\hat{h}_i} r_i^{2\gamma}\Delta t} - \frac{\left(\Delta\hat{h}_i + \mu\hat{h}_i\Delta t\right)^2}{2\sigma_{\hat{h}}^2\Delta t} \right\}. \end{aligned} \tag{3.48}$$

Then according to the Bayes' Theorem, the posterior distribution of each parameter can be obtained from Equation (3.48). Consequently,

$$p((\alpha,\beta)|r,\Theta_{-(\alpha,\beta)}) \propto \prod_{i=0}^{n-1} \exp\left\{ -\frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 e^{\hat{h}_i} r_i^{2\gamma}\Delta t} \right\}, \tag{3.49}$$

$$p(\sigma_r|\Theta_{-\sigma}) \propto \prod_{i=0}^{n-1} \sigma_r^{-1} \exp\left\{ -\frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 e^{\hat{h}_i} r_i^{2\gamma}\Delta t} \right\}, \tag{3.50}$$

$$p(\gamma|\Theta_{-\gamma}) \propto \prod_{i=0}^{n-1} r_i^{-\gamma} \exp\left\{ -\frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 e^{\hat{h}_i} r_i^{2\gamma}\Delta t} \right\}, \tag{3.51}$$

$$p(\mu|\Theta_{-\mu}) \propto \prod_{i=0}^{n-1} \exp\left\{ -\frac{\left(\Delta\hat{h}_i + \mu\hat{h}_i\Delta t\right)^2}{2\sigma_{\hat{h}}^2\Delta t} \right\}, \tag{3.52}$$

$$p(\sigma_h|\Theta_{-\sigma_h}) \propto \prod_{i=0}^{n-1} \sigma_{\hat{h}}^{-1} \exp\left\{ -\frac{\left(\Delta\hat{h}_i + \mu\hat{h}_i\Delta t\right)^2}{2\sigma_{\hat{h}}^2\Delta t} \right\}. \tag{3.53}$$

Details of parameter simulation and the generating latent $\hat{h}$ are presented in following section.

3.3.3 Sample algorithm for parameters of the SVG model

$\hat{h}$

The latent process $\{\hat{h}_i\}$ are sampled from its *full conditional distribution* $p(\hat{h}_i|\hat{h}_{-i})$, which is the distribution of the ith component of $\hat{h}$ conditioning on all the remaining components. It is not difficult to show that $\hat{h}$ is a Markov process for the SVG model from Equation (3.45). Therefore $\hat{h}_i$ depends only on its previous and following state, $\hat{h}_{i-1}$ and $\hat{h}_{i+1}$ respectively. Given $\Delta\hat{h}_i = \hat{h}_{i+1} - \hat{h}_i$, the full conditional distribution of $\hat{h}_i$ will be:

$$p(\hat{h}_i|\hat{h}_{i-1}, \hat{h}_{i+1}) \propto \exp\left\{-\frac{(\Delta\hat{h}_{i-1} + \mu\hat{h}_{i-1}\Delta t)^2 + (\Delta\hat{h}_i + \mu\hat{h}_i\Delta t)^2}{2\sigma_{\hat{h}}^2\Delta t}\right\},$$

which indicates that

$$\hat{h}_i|\hat{h}_{i-1}, \hat{h}_{i+1} \sim N\left(a(\hat{h}_{i-1} + \hat{h}_{i+1}), \frac{\sigma_{\hat{h}}^2\Delta t}{(\mu\Delta t - 1)^2 + 1}\right),$$

where $a = \frac{\mu\Delta t - 1}{(\mu\Delta t - 1)^2 + 1}$. Then the Gibbs sampler can be applied to update the latent process $\{\hat{h}_i\}$. Eraker (2001) suggested the *Accept/Reject Metropolis-Hastings* (AR-MH) algorithm for the simulation of $\hat{h}$. Since the converge rate of Gibbs algorithm is much quicker than that of AR-MH in our case, we will sample $\hat{h}$ from its full conditional distribution directly. The starting point $\hat{h}_0$ is simulated from the distribution $N(0, \frac{\sigma_{\hat{h}}^2}{2\mu})$.

(α, β)

Define vector $y = (y_1, ..., y_n)'$ where $y_i = \Delta r_i e^{-1/2\hat{h}_i} r_i^{-\gamma}\Delta t$ and an $n \times 2$ matrix $X = (x_1, ..., x_i, ..., x_n)'$ where $x_i = (r_i^{-\gamma}e^{-1/2\hat{h}_i}\sqrt{\Delta t}, r_i^{1-\gamma}e^{-1/2\hat{h}_i}\sqrt{\Delta t})'$. Equation (3.49) can be rewritten as:

$$p((\alpha, \beta)|r, \Theta_{-(\alpha,\beta)}) \propto \exp\left\{-\frac{(y'y - 2(\alpha, \beta)X'y + (\alpha, \beta)X'X(\alpha, \beta)')}{2\sigma_r^2}\right\},$$

which is a Normal distribution with mean, $M_1 = (X'X)^{-1}(X'y)$ and variance $S_1^2 = \sigma_r^2(X'X)^{-1}$. As the CKLS model, (α, β) are sampled directly from their normal posterior distribution.

σ_r

Equation (3.50) can be rewritten as:

$$p(\sigma_r|\Theta_{-\sigma_r}) \propto (\sigma_r^{-2})^{n/2} \exp\left\{\sigma_r^{-2} \sum_{i=0}^{n-1} -\frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2r_i^{2\gamma} e^{\hat{h}_i} \Delta t}\right\}.$$

As we have learned from the CKLS model, σ_r^2 follows an Inverse Gamma distribution, $IG(n/2-1, A)$ with $A = \sum_{i=0}^{n-1} \frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2r_i^{2\gamma} e^{\hat{h}_i} \Delta t}$. Again, sample of σ_r is obtained directly from its posterior distribution.

γ

MH algorithm is applied to sample γ since the posterior distribution does not belong to any well-known family. The acceptable probability will be:

$$\alpha(\gamma, \xi_\gamma) = \min\left(1, \frac{P(\xi_\gamma|\Theta_{-\gamma}) f(\gamma|\xi_\gamma)}{P(\gamma|\Theta_{-\gamma}) f(\xi_\gamma|\gamma)}\right) \tag{3.54}$$

where $f(.|.)$ is the proposal distribution. As the CKLS model, we choose the normal distribution with mean γ, and then the ratio $f(\gamma|\xi_\gamma)/f(\xi_\gamma|\gamma) = 1$. Accompanied with Equation (3.51), Equation (3.54) equals:

$$\alpha(\gamma, \xi_\gamma) = \min\left(1, \prod_{i=0}^{n-1} r_i^{(\gamma-\xi_\gamma)} \exp\left\{-A_{i\Delta t}\left(r_i^{-2\xi_\gamma} - r_i^{-2\gamma}\right)\right\}\right),$$

where $A_i = \frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 e^{\hat{h}_i} \Delta t}$.

μ

From Equation (3.52), we find the posterior distribution of μ is a Normal with mean, A and variance, B as following:

$$A = \frac{\sum_{i=0}^{n-1} \hat{h}_i \Delta \hat{h}_i}{\sum_{i=0}^{n-1} \hat{h}_i^2 \Delta t}, \tag{3.55}$$

$$B = \frac{\sigma_{\hat{h}}^2}{\sum_{i=0}^{n-1} \hat{h}_i^2}. \tag{3.56}$$

Simulation of μ will be obtained directly then.

$\sigma_{\hat{h}}$

Obviously, from Equation (3.53), posterior distribution of σ_r will be the same family as σ_h, the Inverse Gamma distribution. σ_r is then simulated directly from $IG(C, D)$ where

$$C = \frac{n}{2} - 1, \tag{3.57}$$

$$D = \sum_{i=0}^{n-1} \frac{(\Delta\hat{h}_i + \mu\hat{h}_i)^2}{2}. \tag{3.58}$$

Now the MCMC sampling algorithm for the SVG model is:

step 0 Initialize Θ at beginning , simulate $\{\hat{h}_i\}$;

step 1 update $\{\hat{h}_i\}$ with Gibbs sampler;

step 2 update (α, β) using Gibbs step from Normal distribution ;

step 3 update σ_r using Gibbs step from Inverse Gamma distribution;

step 4 update γ using M-H step ;

step 5 update μ using Gibbs step from Normal distribution;

step 6 update $\sigma_{\hat{h}}$ using Gibbs step from Inverse Gamma distribution;

step 7 go back to step 2.

Estimation results for the historical data are presented in next chapter.

3.4 MCMC for the SVJ Model

3.4.1 Introduction

The jump process is a point process composed jump times and jump sizes, the latent volatility is therefore determined by these points and then the key role of MCMC of

both SVJ and SVS models will focus on the algorithm of updating the jump point process.

3.4.2 Posterior distributions for the SVJ model

The *Euler* specification of SVJ model is:

$$\Delta r_i = (\alpha + \beta r_i)\Delta t + v_i(\Delta t)^{1/2} r_i^{\gamma} \Delta W_i, \tag{3.59}$$

where $v_i(\Delta t)$ is the integrated volatility defined in section 2.5 from the ith time unit to the $i+1$th time unit:

$$v_i(\Delta t) = v_0((i+1)\Delta t) - v_0(i\Delta t). \tag{3.60}$$

From both Equation (2.30) and (2.31), we can get following:

$$\begin{aligned} v_0(i\Delta t) &= \frac{1}{\mu} \sum_{j=1}^{N_0(i\Delta t)} \varepsilon_j - \frac{1}{\mu}(h_i - h_0), \\ &= \frac{1}{\mu} \left(h_0 \left(1 - e^{-\mu i \Delta t}\right) + \sum_{j=1}^{N_0(i\Delta t)} \varepsilon_j \left(1 - e^{-\mu(i\Delta t - c_j)}\right) \right). \end{aligned} \tag{3.61}$$

where $N_0(i\Delta t)$ is the number of jumps happening from the beginning till the ith time unit, $\{c_j\}$ is the jump time and $\{\varepsilon_j\}$ is the jump size from $Exponential(\sigma_h^{-1})$. The integrated volatility is therefore fully determined by the marked Poisson process: $\Psi(i) = \{(c_j, \varepsilon_j)\}^{N_0(i\Delta t)}$ on the space $(0, i\Delta t) \times (0, \infty)$.

In order to increase the convergence rate of MCMC, we will follow the the so-called *Non-centred parameterizations* (NCP) for the SVJ model developed by Papaspiliopoulos (2003). The NCP is designed to decrease the dependency between the parameters and the latent process. In particular, we construct a new marked Poisson process $\tilde{\Psi} = \{(c_j, k_j, \tilde{\varepsilon}_j)\}$ on higher dimensional space $[0, T] \times (0, \infty) \times (0, \infty)$ than Ψ. $\{c_j\}$ still presents jump time, $\{k_j\}$ is a positive random process uniformly distributed in $(0, \infty)$, and $\{\tilde{\varepsilon}_j\}$ is a process independent of both $\{c_j\}$ and $\{k_j\}$. $\tilde{\varepsilon}_j$ follows exponential distribution with intensity parameter equal to 1 for all j. Consequently, $\tilde{\Psi}$ is independent of other parameters. Ψ can be obtained from $\tilde{\Psi}$ by the following steps:

1. Pick up the points from $\tilde{\Psi}$ with $k_j < \lambda$.

2. Project them to $[0, T] \times (0, \infty)$ and obtain process $\{(c_j, \tilde{\varepsilon}_j)\}$.

3. Transform $\{(c_j, \tilde{\varepsilon}_j)\}$ to $\{(c_j, \varepsilon_j)\}$ where $\varepsilon_j = \tilde{\varepsilon}_j / \sigma_h$.

More details about construction of $\tilde{\Psi}$ and transformation from $\tilde{\Psi}$ to Ψ can be found in Papaspiliopoulos (2003). Another transformation considered by him is the one from h_0 to $\tilde{h}_0 = h_0/\sigma_h$ which will result a conditional Gibbs step for σ_h and also make σ_h a priori independent of ν. The modified items structure of the SVJ model is shown in Figure 3.9.

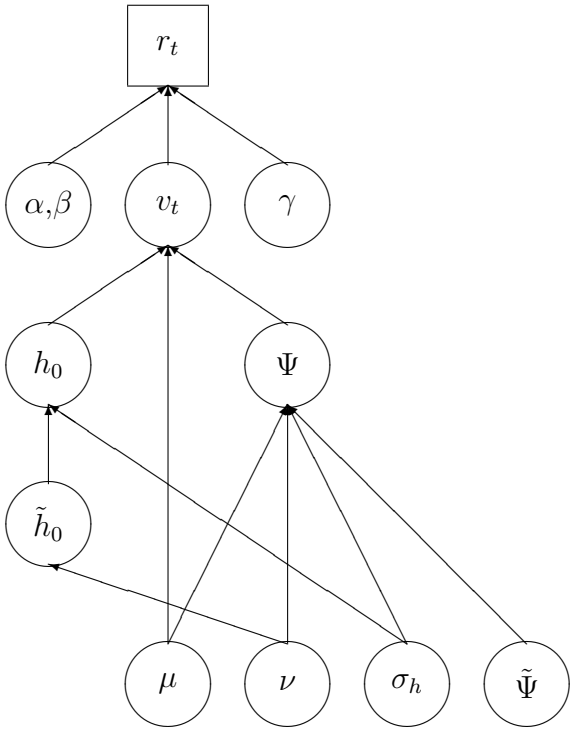

Figure 3.9: Graphical model of the non-centered parameterizations for the SVJ model

Then the posterior distribution of unknown parameters and latent data can be written as:

$$p(\Theta, \tilde{h}_0, \tilde{\Psi}|r) \propto L(r|\Theta, \tilde{h}_0, \tilde{\Psi}) p(\tilde{h}_0|\Theta) p(\tilde{\Psi}) p(\Theta) \tag{3.62}$$

where $\Theta = (\alpha, \beta, \gamma, \mu, \nu, \sigma_h)'$. The likelihood function of r_t has the following form:

$$L(r|\Theta, \tilde{h}_0, \tilde{\Psi}) \propto \prod_{i=0}^{n-1} \sigma_h^{-1/2} \tilde{v}_i(\Delta t)^{-1/2} r_i^{-\gamma} \Delta t^{-1/2} \exp\left\{ -\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\sigma_h \tilde{v}_i(\Delta t) r_i^{2\gamma} \Delta t} \right\},$$

and the conditional distribution of $\tilde{h}_0$ is a Gamma distribution

$$p(\tilde{h}_0|\Theta) = \frac{1}{\Gamma(\nu)} \tilde{h}_0^{\nu-1} \exp(-\tilde{h}_0).$$

Same as previous sections, we will use the independent uninformative prior distribution of parameters, then Equation (3.62) will have the form:

$$p(\Theta, \tilde{h}_0, \tilde{\Psi}|r) \propto \frac{1}{\Gamma(\nu)} \tilde{h}_0^{\nu-1} \sigma_h^{-n/2} A_1 A_2 \exp\left\{ -\sigma_h^{-1} A_3 - \tilde{h}_0 \right\} \tag{3.63}$$

where

$$\begin{aligned}
A_1 &= \prod_{i=0}^{n-1} \frac{1}{\mu} \left(\sum_{j=1}^{N_i(\Delta t)} \tilde{\varepsilon}_j - (e^{-\mu(i+1)\Delta t} - e^{-\mu i \Delta t})(\tilde{h}_0 + \sum_{j=1}^{N_0(i\Delta t)} e^{\mu c_j} \tilde{\varepsilon}_j) - \sum_{j=1}^{N_i(\Delta t)} e^{-\mu((i+1)\Delta t - c_j)} \tilde{\varepsilon}_j \right), \\
A_2 &= \prod_{i=0}^{n-1} r_i^{-\gamma}, \\
A_3 &= \sum_{i=0}^{n-1} \frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\tilde{v}_i(\Delta t) r_i^{2\gamma} \Delta t},
\end{aligned}$$

and $N_i((j-i)\Delta t)$ is the number of jumps on the interval from the ith time unit to jth time unit.

Consequently,

$$p((\alpha, \beta)|r, \Theta_{-(\alpha,\beta)}) \propto \prod_{i=0}^{n-1} \exp\left\{ -\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\sigma_h \tilde{v}_i(\Delta t) r_i^{2\gamma} \Delta t} \right\}, \tag{3.64}$$

$$p(\gamma|\Theta_{-\gamma}) \propto \prod_{i=0}^{n-1} r_i^{-\gamma} \exp\left\{ -\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\sigma_h \tilde{v}_i(\Delta t) r_i^{2\gamma} \Delta t} \right\}, \tag{3.65}$$

$$p((\mu, \nu)|\Theta_{-(\mu,\nu)}) \propto \frac{1}{\Gamma(\nu)} \tilde{h}_0^{\nu-1} A_1 \exp\left\{ -\sigma_h^{-1} A_3 \right\}, \tag{3.66}$$

$$p(\sigma_h|\Theta_{-\sigma_h}) \propto \sigma_h^{-n/2} \exp\left\{ -\sigma_h^{-1} A_3 \right\}, \tag{3.67}$$

$$p(\tilde{h}_0|\Theta_{-\tilde{h}_0}) \propto \tilde{h}_0^{\nu-1} A_1 \exp\left\{ -\tilde{h}_0 \right\}. \tag{3.68}$$

where $\Gamma()$ is Gamma function.

3.4.3 Simulation of the jump process

In this section, we will present the algorithm of both simulation and updating, Ψ the marked Poisson process which is composed the latent volatility of both SVJ and SVS model. The updating algorithm is mainly based on the work of Papaspiliopoulos *et al.* (2004).

Let $\Psi = \{(c_i, \varepsilon_i)\}_{i=1}^{N}$ be a compound Poisson process where c_i is the jump time and ε_i is the jump size at c_i. Its trajectory in time duration $(0, t)$is given by:

$$Y(t) = \sum_{i=1}^{N(0,t)} \varepsilon_i$$

where $N(0, t)$ is the number of jumps in $(0, t)$ which follows a Poisson distribution with intensity λt since λ is the jump rate in $(0, 1)$. Figure 3.10 provides two examples of the Poisson marked process with same Poisson rate λ but different jump size distribution.

The sample of Ψ in NCP structure is obtained from $\tilde{\Psi}$ following the steps described in previous section. As we have defined in section 3.4.2, $\tilde{\Psi} = \{(c_i, k_i, \tilde{\varepsilon}_i)\}$ is a marked Poisson process on a higher dimensional space $(0, t) \times (0, \infty) \times (0, \infty)$.

Simulation of such process can be obtained by these steps. First, simulate $N(0, t)$ from $Poi(\tilde{\lambda}t)$, $\tilde{\lambda} > \lambda$ which is the intensity parameter of Ψ. Generate $N(0, t)$ independent random variables from the uniform distribution on the interval $(0, t)$, and the jump times $\{c_i\}$ are composed by the ordered variables. The same number of m_i are simulated uniformly from the interval $(0, a)$ where a is a positive real number and $a > \lambda$. Finally, the $N(0, t)$ jump sizes are sampled independently, and in our case, the distribution is exponential distribution with unit intensity.

It will be very inefficient if we regenerate Ψ at each MCMC iteration. The methodology we use here to update the point process is based on the work of Papaspiliopoulos *et al.* (2004). We randomly choose between two kinds of move at each iteration: the *Birth & Death* move and the *Displacement* move. The first one is to randomly add

one point in Ψ or remove one point from it, and the second one is to choose one point randomly in Ψ and replace it with new one. Details of these methodologies are presented in the Appendix B and the Appendix C respectively.

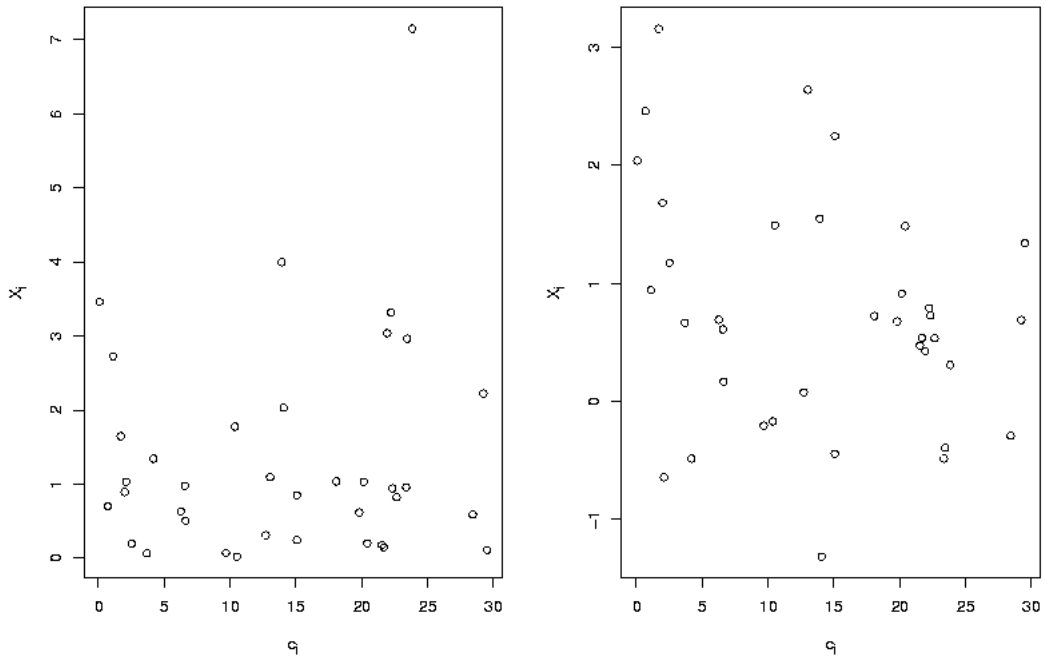

Figure 3.10: The marked Poisson process Ψ: left: the jump size follows an exponential distribution with rate $\theta = 1$; right: the jump size follows a Normal distribution with both mean and variance equal to 1

3.4.4 Sample algorithm for the parameters of the SVJ model

(α, β)

Define vector $y = (y_1, ..., y_n)'$ where $y_i = \Delta r_i v_i (\Delta t)^{-1/2} r_i^{-\gamma} \Delta t$ and an $n \times 2$ matrix $X = (x_1, ..., x_n)'$ where $x_i = (r_i^{-\gamma}\sqrt{\Delta t}, r_i^{1-\gamma} v_i (\Delta t)^{-1/2}\sqrt{\Delta t})'$. Equation (3.64) can be

rewritten as:

$$p((\alpha,\beta)|r,\Theta_{-(\alpha,\beta)}) \propto \exp\left\{-\frac{(y'y - 2(\alpha,\beta)X'y + (\alpha,\beta)X'X(\alpha,\beta)')}{2\sigma_r^2}\right\},$$

which is a Normal distribution with mean, $M_1 = (X'X)^{-1}(X'y)$ and variance $S_1^2 = \sigma_r^2(X'X)^{-1}$. As the CKLS model, ($\alpha$, β) are sampled directly from their normal posterior distribution.

γ

MH algorithm is applied to sample γ since the posterior distribution does not belong to any well-known family. The acceptable probability will be:

$$\alpha(\gamma,\xi_\gamma) = \min\left(1, \frac{p(\xi_\gamma|\Theta_{-\gamma})f(\gamma|\xi_\gamma)}{p(\gamma|\Theta_{-\gamma})f(\xi_\gamma|\gamma)}\right) \tag{3.69}$$

where $f(.|.)$ is the proposal distribution. As the CKLS model, we choose the normal distribution with mean γ, and then the ratio $f(\gamma|\xi_\gamma)/f(\xi_\gamma|\gamma) = 1$. Accompanied with Equation (3.65), Equation (3.69) equals:

$$\alpha(\gamma,\xi_\gamma) = \min\left(1, \prod_{i=0}^{n-1} r_i^{(\gamma-\xi_\gamma)} \exp\left\{-A_i\left(r_i^{-2\xi_\gamma} - r_i^{-2\gamma}\right)\right\}\right),$$

where $A_i = \frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 v_i(\Delta t)\Delta t}$.

(μ, ν)

We will use a random-walk Metropolis-Hastings algorithm to update (μ, ν). Let $\Theta_\lambda = (\mu,\nu)$ and $\xi_\lambda = (\xi_\mu,\xi_\nu)$ be the candidate vector which are generated independently from Normal distribution. From Equation (3.66), the acceptance probability is

$$\alpha(\Theta_\lambda,\xi_\lambda) = \min\left(1, \tilde{h}_0^{\xi_\nu-\nu}\frac{\Gamma(\nu)A_1'}{\Gamma(\xi_\nu)A_1}\exp\left\{-\sigma_h^{-1}(A_3' - A_3)\right\}\right),$$

where A' is calculated not only with the candidate values but also with the updating point process Ψ'. With the candidates ξ_μ and ξ_ν, we get the new $\lambda' = \xi_\nu\xi_\mu$. Ψ will be updated with this new intensity parameter to Ψ' which are used to calculate the integrated volatility.

σ_h

From Equation (3.67), the posterior law for σ_h^{-1} is a Gamma distribution with parameter $n/2+1$ and A_3. As we know, the Gamma distribution will be close to Normal distribution when parameter $n/2+1$ is sufficiently large. Therefore, we can simulate σ_h^{-1} directly from Normal distribution with mean $(n/2+1)/A_3$, variance $(n/2+1)/A_3^2$ and the sample size $n = 2242$.

$\tilde{h}_0$

$\tilde{h}_0$ will also be updated using a random-walk Metropolis-Hastings algorithm. The acceptance probability for $\tilde{h}_0$ is

$$\alpha(\tilde{h}_0, \xi_h) = \min\left(\left(\frac{\xi_{\tilde{h}_0}}{\tilde{h}_0}\right)^{\nu-1} \frac{A_1'}{A_1} \exp\left\{-(\xi_{\tilde{h}_0} - \tilde{h}_0)\right\}\right),$$

where $\xi_{\tilde{h}_0}$ is proposed from a Normal distribution.

The MCMC algorithm to sample parameters of the SVJ model is:

step 0 Initialize all unknown parameters, simulate $\tilde{\Psi}$;

step 1 update (μ, ν) using M-H step;

step 2 update σ_h using Gibbs step from inverse Gamma posterior distribution;

step 3 transformation: $\tilde{h}_0 \to h_0$ and $\tilde{\Psi} \to \Psi$;

step 4 update h_0 using M-H step;

step 5 update Ψ using M-H step;

step 6 update (α, β) using Gibbs step from joint Normal posterior distribution;

step 7 update γ using M-H step;

step 8 transformation: $h_0 \to \tilde{h}_0$ and $\Psi \to \tilde{\Psi}$;

step 9 go back to step 1.

Estimation results for the historical data are presented in next chapter.

3.4.5 MCMC for the SVS model

3.4.6 Posterior distributions for the SVS model

The *Euler* specification of SVS model is:

$$\Delta r_i = (\alpha + \beta r_i)\Delta t + v_i(\Delta t)^{1/2} r_i^{\gamma} \Delta W_i, \tag{3.70}$$

where $v_i(\Delta t)$ is the integrated volatility defined as below:

$$v_i(\Delta t) = v_{1,i}(\Delta t)) + v_{2,i}(\Delta t), \tag{3.71}$$

and $v_{m,i}(\Delta t)$, for $m = 1, 2$ is

$$v_{m,i}(\Delta t) = \frac{1}{\mu_m}\left(h_{m,i}\left(1 - e^{-\mu_m \Delta t}\right) + \sum_{j=1}^{N_{m,i}(\Delta t)} \varepsilon_{m,j}\left(1 - e^{-\mu_m((i+1)\Delta t - c_{m,j})}\right)\right), \tag{3.72}$$

$\Psi_m = \{(c_{m,j}, \varepsilon_{m,j})\}$ is the marked Poisson process for the jth part of latent volatility, and $\Psi = \Psi_1 \cup \Psi_2$.

For the SVS model, we will also use the NCP structure described in the SVJ model. First, we construct a higher dimensional point process $\tilde{\Psi} = \{(c_j, k_j, \tilde{\varepsilon}_j)\}$ on the space $(0, t) \times (-\infty, +\infty) \times (0, +\infty)$ where $k - j$ is a random variable uniformly distributed in $(-\infty, +\infty)$. $\tilde{\Psi} = \tilde{\Psi}_1 \cup \tilde{\Psi}_2$ where $\tilde{\Psi}_1$ contains all the points with positive k_j and $\tilde{\Psi}_2$ contains the others. Then transformation from $\tilde{\Psi}_m$ to Ψ_m is similarly as the steps of the SVJ model by first picking up the points with $|k_{m,j}| < \lambda_m$ for $m = 1, 2$ and then projecting such points to the 2-dimensional space $(0, t) \times (0, +\infty)$.

From the graph model presented in Figure 3.11, the joint posterior distribution is:

$$\begin{aligned} p(\Theta, \tilde{\Psi}_1, \tilde{\Psi}_2, \tilde{h}_{1,0}, \tilde{h}_{2,0}|r) \quad &\propto \quad L(r|\Theta, \tilde{\Psi}_1, \tilde{\Psi}_2, \tilde{h}_{1,0}, \tilde{h}_{2,0}) \\ &\quad p(\tilde{h}_{1,0}|\Theta) p(\tilde{h}_{2,0}|\Theta) p(\tilde{\Psi}_1) p(\tilde{\Psi}_2) p(\Theta) \end{aligned} \tag{3.73}$$

where the parameter vector $\Theta = (\alpha, \beta, \gamma, \nu_1, \mu_2, \nu_2, \mu_2, \sigma_h)$ and $\tilde{h}_{1,0} = h_{1,0}/\sigma_h$, $\tilde{h}_{2,0} = h_{2,0}/\sigma_h$ respectively. The likelihood function of r is

$$\begin{aligned} p(r|\Theta, \tilde{h}_{1,0}, \tilde{h}_{2,0}, \tilde{\Psi}) \quad &\propto \quad \prod_{i=0}^{n-1} \sigma_h^{-1/2} \tilde{v}_i(\Delta t)^{-1/2} r_i^{-\gamma} \Delta t^{-1/2} \\ &\quad \exp\left\{ -\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\sigma_h \tilde{v}_i(\Delta t) r_i^{2\gamma} \Delta t} \right\}, \end{aligned}$$

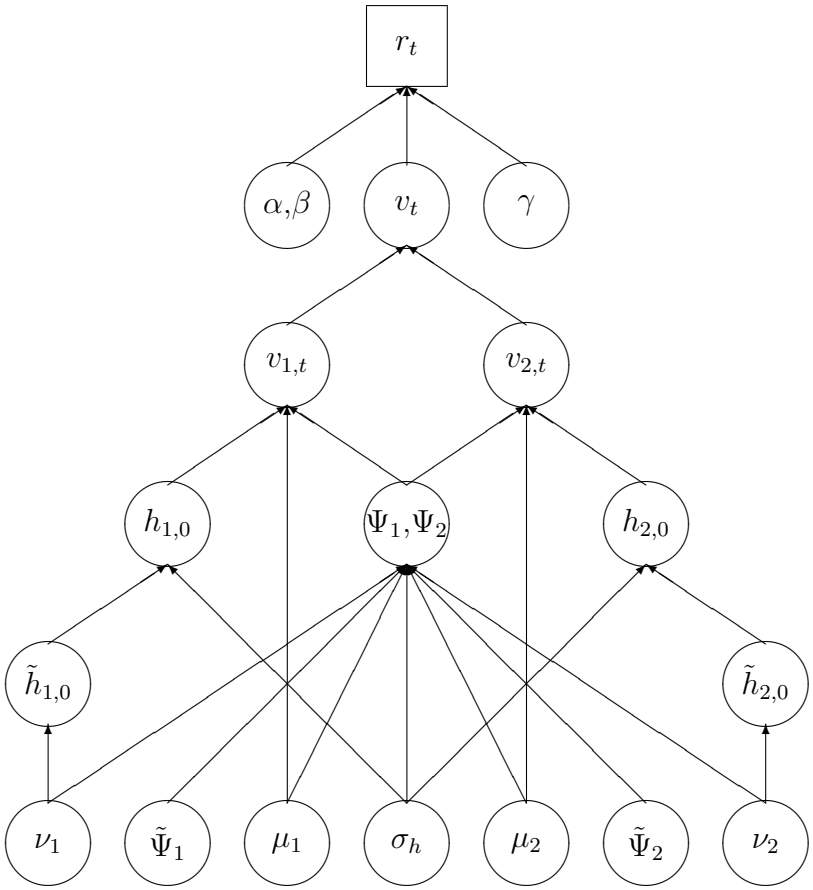

Figure 3.11: Graphical model of the non-centered parameterizations of the SVS model

and the conditional distribution of $\tilde{h}_{m,0}$ is a Gamma distribution for each $m = 1, 2$

$$p(\tilde{h}_{m,0}|\Theta) = \frac{1}{\Gamma(\nu_m)} \tilde{h}_{m,0}^{\nu_m - 1} \exp(-\tilde{h}_{m,0}).$$

Using the independent uninformative prior distribution of parameters, then Equation (3.73) will have the form:

$$\begin{aligned} p(\Theta, \tilde{h}_{1,0}, \tilde{h}_{2,0}, \tilde{\Psi}|r) \quad &\propto \quad \frac{1}{\Gamma(\nu_1)\Gamma(\nu_2)} \tilde{h}_{1,0}^{\nu_1 - 1} \tilde{h}_{2,0}^{\nu_2 - 1} \sigma_h^{-n/2} A_1 A_2 \\ &\qquad \exp\left\{-\sigma_h^{-1} A_3 - \tilde{h}_{1,0} - \tilde{h}_{2,0}\right\} \end{aligned} \tag{3.74}$$

where

$$\begin{aligned} A_1 &= \prod_{i=0}^{n-1} \tilde{v}_i (\Delta t)^{-1/2}, \\ A_2 &= \prod_{i=0}^{n-1} r_i^{-\gamma}, \\ A_3 &= \sum_{i=0}^{n-1} \frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\tilde{v}_i(\Delta t) r_i^{2\gamma} \Delta t}. \end{aligned}$$

The integrated volatility given $\tilde{\Psi}$ is the superposition of two independent OU process $\tilde{v}_1$ and $\tilde{v}_2$ which can be calculated as

$$\begin{aligned} \tilde{v}_{m,i}(\Delta t) \quad &= \quad \frac{1}{\mu_m} \sum_{j=1}^{N_{m,i}(\Delta t)} \tilde{\varepsilon}_{m,j} - \frac{1}{\mu_m} (e^{-\mu_m (i+1)\Delta t} - e^{-\mu_m i \Delta t})(\tilde{h}_{m,0} + \sum_{j=1}^{N_{m,0}(i\Delta t)} e^{\mu_m c_{m,j}} \tilde{\varepsilon}_{m,j}) \\ &\qquad - \frac{1}{\mu_m} \sum_{j=1}^{N_{m,i}(\Delta t)} e^{-\mu_m ((i+1)\Delta t - c_{m,j})} \tilde{\varepsilon}_{m,j} \end{aligned} \tag{3.75}$$

where $N_{m,i}((j-i)\Delta t)$ is the number of jumps from the ith time unit to the jth time unit for the mth point process.

Consequently, the posterior distributions for each parameter of the SVS model are

$$p((\alpha, \beta)|r, \Theta_{-(\alpha,\beta)}) \propto \prod_{i=0}^{n-1} \exp\left\{-\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\sigma_h \tilde{v}_i(\Delta t) r_i^{2\gamma} \Delta t}\right\}, \tag{3.76}$$

$$p(\gamma|\Theta_{-\gamma}) \propto \prod_{i=1}^{n} r_i^{-\gamma} \exp\left\{-\frac{(\Delta r_i - (\alpha + \beta r_i)\Delta t)^2}{2\sigma_h \tilde{v}_i(\Delta t) r_i^{2\gamma} \Delta t}\right\}, \tag{3.77}$$

$$p((\mu_1, \nu_1, \mu_2, \nu_2)|\Theta_{-(\mu_1,\nu_1,\mu_2,\nu_2)}) \propto \frac{1}{\Gamma(\nu_1)\Gamma(\nu_2)} \tilde{h}_{1,0}^{\nu_1 - 1} \tilde{h}_{2,0}^{\nu_2 - 1} A_1 e^{(-\sigma_h^{-1} A_3)}, \tag{3.78}$$

$$p(\sigma_h|\Theta_{-\sigma_h}) \propto \sigma_h^{-n/2} e^{(-\sigma_h^{-1} A_3)}, \tag{3.79}$$

$$p(\tilde{h}_{m,0}|\Theta_{-\tilde{h}_{k,0}}) \propto \tilde{h}_{m,0}^{\nu_m - 1} A_1^{(-\tilde{h}_{m,0})}, m = 1, 2. \tag{3.80}$$

3.4.7 Sample algorithm of the parameters of the SVS model

$(\alpha,\ \beta)$

Define vector $y = (y_1, ..., y_n)'$ where $y_i = \Delta r_i v_i (\Delta t)^{-1/2} r_i^{-\gamma} \Delta t$ and an $n \times 2$ matrix $X = (x_1, ..., x_n)'$ where $x_i = (r_i^{-\gamma}\sqrt{\Delta t}, r_i^{1-\gamma} v_i (\Delta t)^{-1/2}\sqrt{\Delta t})'$. Equation (3.76) can be rewritten as:

$$p((\alpha,\beta)|r, \Theta_{-(\alpha,\beta)}) \propto \exp\left\{-\frac{(y'y - 2(\alpha,\beta)X'y + (\alpha,\beta)X'X(\alpha,\beta)')}{2\sigma_r^2}\right\},$$

which is a Normal distribution with mean, $M_1 = (X'X)^{-1}(X'y)$ and variance $S_1^2 = \sigma_r^2 (X'X)^{-1}$. As the CKLS model, $(\alpha,\ \beta)$ are sampled directly from their normal posterior distribution.

γ

MH algorithm is applied to sample γ since the posterior distribution does not belong to any well-known family. The acceptance probability will be:

$$\alpha(\gamma, \xi_\gamma) = \min\left(1, \frac{P(\xi_\gamma|\Theta_{-\gamma}) f(\gamma|\xi_\gamma)}{P(\gamma|\Theta_{-\gamma}) f(\xi_\gamma|\gamma)}\right) \tag{3.81}$$

where $f(.|.)$ is the proposal distribution. As the CKLS model, we choose the normal distribution with mean γ, and then the ratio $f(\gamma|\xi_\gamma)/f(\xi_\gamma|\gamma) = 1$. Accompanied with Equation (3.77), Equation (3.69) equals:

$$\alpha(\gamma, \xi_\gamma) = \min\left(1, \prod_{i=1}^{n} r_i^{(\gamma-\xi_\gamma)} \exp\left\{-A_i\left(r_i^{-2\xi_\gamma} - r_i^{-2\gamma}\right)\right\}\right),$$

where $A_i = \frac{(\Delta r_i - (\alpha+\beta r_i)\Delta t)^2}{2\sigma_r^2 v_i (\Delta t)\Delta t}$.

$(\mu_1,\ \nu_1, \mu_2,\ \nu_2)$

Since μ_m and ν_m are the parameters determining the point process Ψ_m, the changing of such parameters will generate different processes, then we have to update them at

the same time. We will use a random-walk Metropolis-Hastings algorithm to update $(\mu_1, \nu_1, \mu_2, \nu_2)$. Let $\Theta_\lambda = (\mu_1, \nu_1, \mu_2, \nu_2)$ and $\xi_\lambda = (\xi_{\mu_1}, \xi_{\nu_1}, \xi_{\mu_2}, \xi_{\nu_2})$ be the candidate vector which are generated independently from Normal distribution. From Equation (3.78), the acceptance probability is

$$\alpha(\Theta_\lambda, \xi_\lambda) = \min\left(1, \tilde{h}_{1,0}^{\xi_{\nu_1}-\nu_1}\tilde{h}_{2,0}^{\xi_{\nu_2}-\nu_2}\frac{\Gamma(\nu_1)\Gamma(\nu_2)A_1^{'}}{\Gamma(\xi_{\nu_1})\Gamma(\xi_{\nu_2})A_1}\exp\left\{-\sigma_h^{-1}(A_3^{'}-A_3)\right\}\right),$$

where A' is calculated not only with the candidate values but also with the updating point process Ψ'. With the candidates ξ_μ and ξ_ν, we get the new $\lambda' = \xi_\nu\xi_\mu$. Ψ will be updated with this new intensity parameter to Ψ' which are used to calculate the integrated volatility.

σ_h

Similarly as the SVJ model, from Equation (3.79), the posterior law for σ_h^{-1} is a Gamma distribution with parameter $n/2+1$ and A_3. Since $n/2+1 = 1122$ is sufficiently large, we can also simulate σ_h^{-1} directly from Normal distribution with mean $(n/2+1)/A_3$ and variance $(n/2+1)/A_3^2$.

$\tilde{h}_{m,0}$

$\tilde{h}_{m,0}$ will also be updated using a random-walk Metropolis-Hastings algorithm. The acceptance probability for $\tilde{h}_{m,0}$ is

$$\alpha(\tilde{h}_{m,0}, \xi_h) = \min\left(\left(\frac{\xi_{h_k}}{\tilde{h}_{m,0}}\right)^{\nu_m-1}\frac{A_1^{'}}{A_1}\exp\left\{-(\xi_{h_k}-\tilde{h}_{m,0})\right\}\right),$$

where ξ_{h_k} is proposed from a Normal distribution for the kth part.

The MCMC algorithm for the SVS model sampling is:

step 0 Initialize all unknown parameters, simulate $\tilde{\Psi}$;

step 1 update $(\mu_1, \nu_1, \mu_2, \nu_2)$ using M-H step;

step 2 update σ_h using Gibbs step from inverse Gamma posterior distribution;

step 3 transformation: $\tilde{h}_{m,0} \to h_{m,0}$ and $\tilde{\Psi}_m \to \Psi_m$ for $m = 1, 2$;

step 4 update $h_{m,0}$ using M-H step for $m = 1, 2$;

step 5 update Ψ_m using M-H step with two kinds of move for $m = 1, 2$;

step 6 update (α, β) using Gibbs step from joint Normal posterior distribution;

step 7 update γ using M-H step;

step 8 transformation: $h_{m,0} \to \tilde{h}_{m,0}$ and $\Psi_m \to \tilde{\Psi}_m$ for $m = 1, 2$;

step 9 go back to step 1.

Estimation results for the historical data are presented in next chapter.

3.5 Simulation Tests

Before we apply these algorithms to the short rate models with historical data, we will test them with some artificial data to assess their performance. The sample we use is simulated from candidate models given predetermined value of parameters which are listed in Table 3.2 to Table 3.5 as the *True* values besides corresponding parameters. The size of sample for each model is 2242. The time step, Δt is set at 1 week: just as the historical data we will analysis in later chapter.

Results of the simulation test for every parameter are presented from Table 3.2 to Table 3.5. We list true values, posterior means, standard deviations and 95% confidence intervals for every model's parameters. These are the results calculated from the samples composed from 6 million iterations with ignoring the first 1 million as burn-in and retaining every 100th variable. Figure 3.12 to 3.18 present the iteration paths, autocorrelations and posterior densities of the parameters for each model. Reports of each model are given in following subsections.

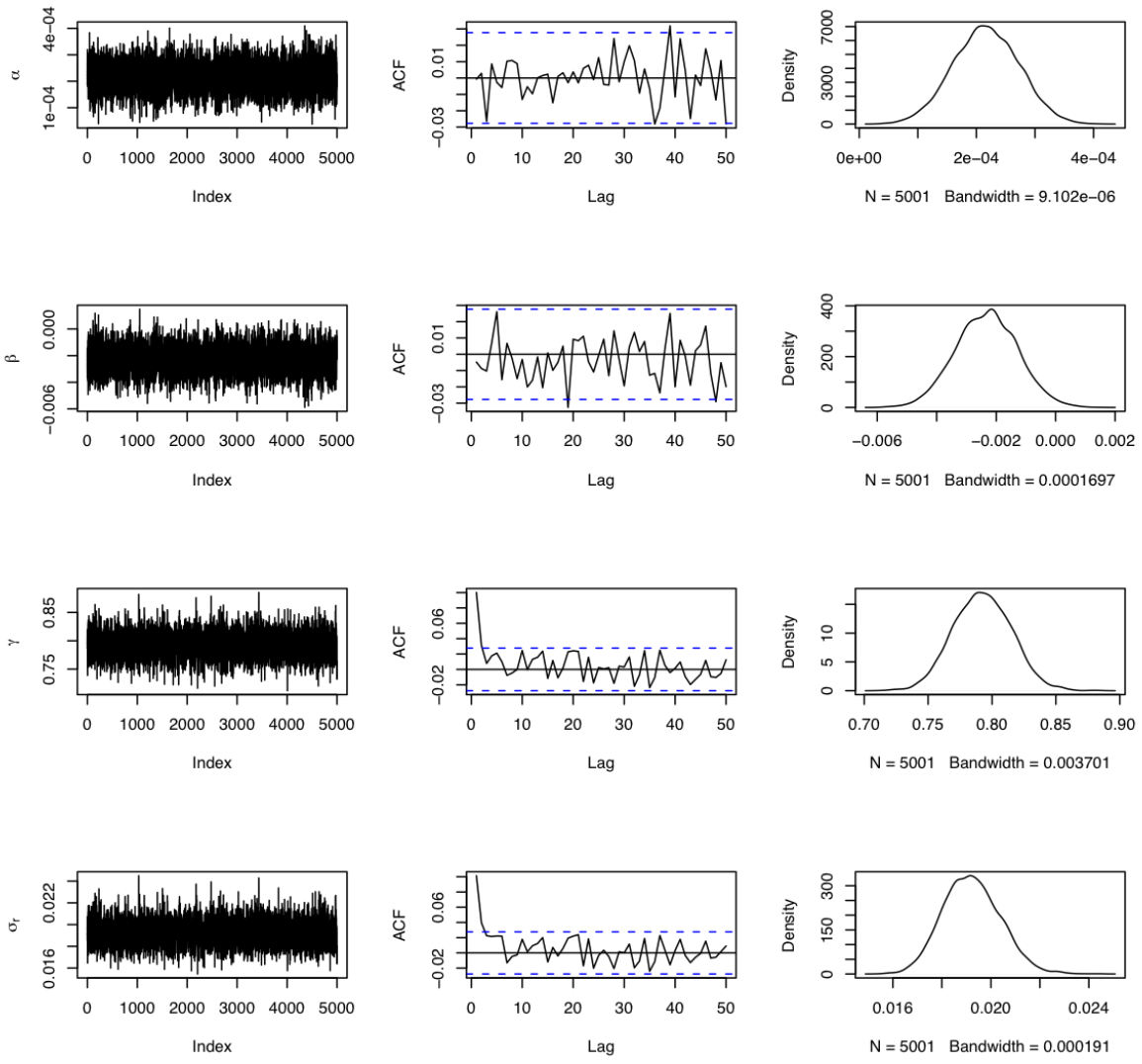

Figure 3.12: The simulation test of CKLS model: left column: the sample paths; middle column: the first 50th autocorrelations (excluding the first); right column: posterior densities.

Table 3.2: Simulation Test for the CKLS Model

Parameter	**True**	**Mean**	**Std.**	95% **C.I.**
α ($\times 10^{-4}$)	2.00	2.16	0.56	(1.07, 3.24)
β ($\times 10^{-3}$)	-3.00	-2.32	1.04	(-4.34, -0.29)
γ ($\times 10^{-1}$)	8.00	7.92	0.23	(7.49, 8.36)
σ_r ($\times 10^{-2}$)	2.00	1.92	0.12	(1.71, 2.16)

Table 3.3: Simulation Test for the SVG Model

Parameter	**True**	**Mean**	**Std.**	95% **C.I.**
α ($\times 10^{-4}$)	2.00	1.98	0.09	(1.79, 2.15)
β ($\times 10^{-3}$)	-3.00	-2.43	0.81	(-4.00, -0.828)
γ ($\times 10^{-1}$)	8.00	8.35	0.95	(6.50, 10.29)
σ_r ($\times 10^{-2}$)	3.00	3.02	0.80	(1.78, 4.91)
σ_h ($\times 10^{-1}$)	6.00	7.10	0.54	(6.08, 8.21)
μ_1 ($\times 10^{-2}$)	7.00	8.96	1.39	(6.40, 11.83)

3.5.1 Simulation test for the CKLS model

Table 3.2 presents the posterior mean, standard deviation and the 95% confidence intervals of each parameter of the CKLS model. As we can see, the MCMC algorithm gives us an acceptable results as all the true values fall into the confidence interval and the estimates are very close to them.

We plot the iteration paths of each parameter in Figure 3.12, along with these, both the first fiftieth autocorrelation and historical posterior densities. The MH algorithm works well to each parameter. As we can observe, the sample converges quickly to its stationary state and the dependence drop dramatically to zero.

3.5.2 Simulation tests for the SVG model

Simulation results are summarized in Table 3.3. As the CKLS model, the parameters of the short rate part are close to their true values. The biases are bigger

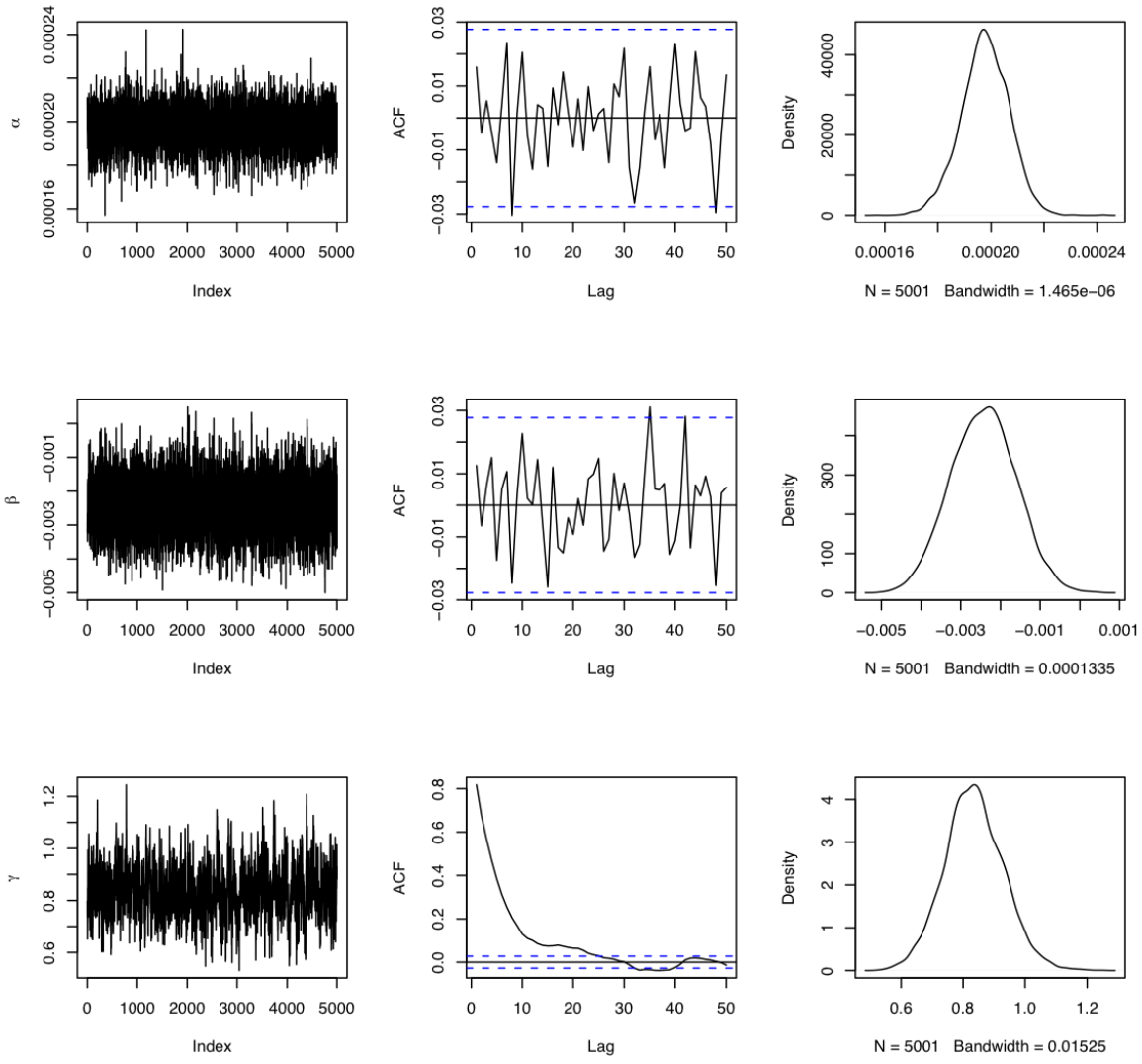

Figure 3.13: The simulation test of SVG model–part 1: left column: the sample paths; middle column: the first 50th autocorrelations (excluding the first); right column: posterior densities.

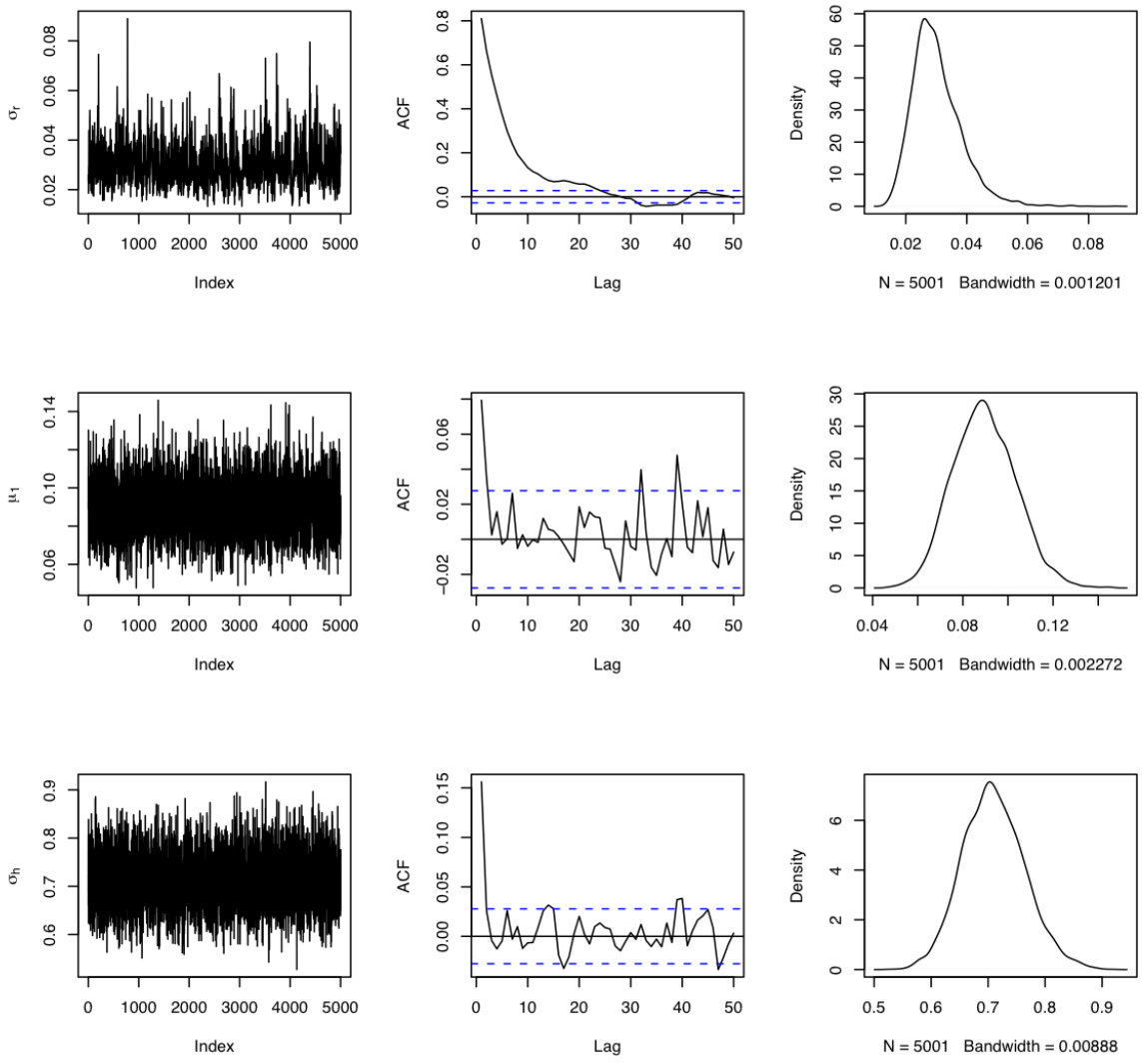

Figure 3.14: The simulation test of SVG model–part 2: left column: the sample paths; middle column: the first 50th autocorrelations (excluding the first); right column: posterior densities.

Table 3.4: Simulation Test for the SVJ Model

Parameter	**True**	**Mean**	**Std.**	95% **C.I.**
α ($\times 10^{-4}$)	2.00	2.16	0.65	(0.94, 3.45)
β ($\times 10^{-3}$)	-3.00	-2.99	0.97	(-4.89, -1.15)
γ ($\times 10^{-1}$)	8.00	8.46	0.50	(7.29, 9.41)
σ_h ($\times 10^{-4}$)	2.00	2.07	0.34	(1.57, 2.87)
μ ($\times 10^{-2}$)	5.00	5.21	0.51	(4.34, 6.34)
ν	1.50	1.46	0.15	(1.18, 1.79)

in volatility part but both of their confidence intervals include the predetermined values.

Another evidence of the fitness of algorithm is shown in both Figure 3.13 and Figure 3.14. Parameter iterations of both part converge quickly. Two parameters are affected by the introduce of latent volatility process. Both γ and σ_r are not easy to be estimated. It takes some steps to make the autocorrelation around zero, and the sample paths of these two are not stable as others. This is also reflected by the increasing of their standard deviations comparing with the CKLS model.

3.5.3 Simulation tests for the SVJ model

Table 3.4 presents the test results of the SVJ model. All the true parameter values fall into the corresponding confidence interval. Theestimations are close to the true values, except γ which is biased more than the others.

When we check the sample paths plotted in Figure 3.15 and Figure 3.16, both γ and parameters of the latent process converge slowly to the stationary state. For each of these parameters, the autocorrelation functions decrease slowly which indicates that not only is a long sample needed but also we should use a long distance between the recorded data. Although the higher dependence reduces the efficiency of the MCMC algorithm, it still works well, with the acceptable results compared with the

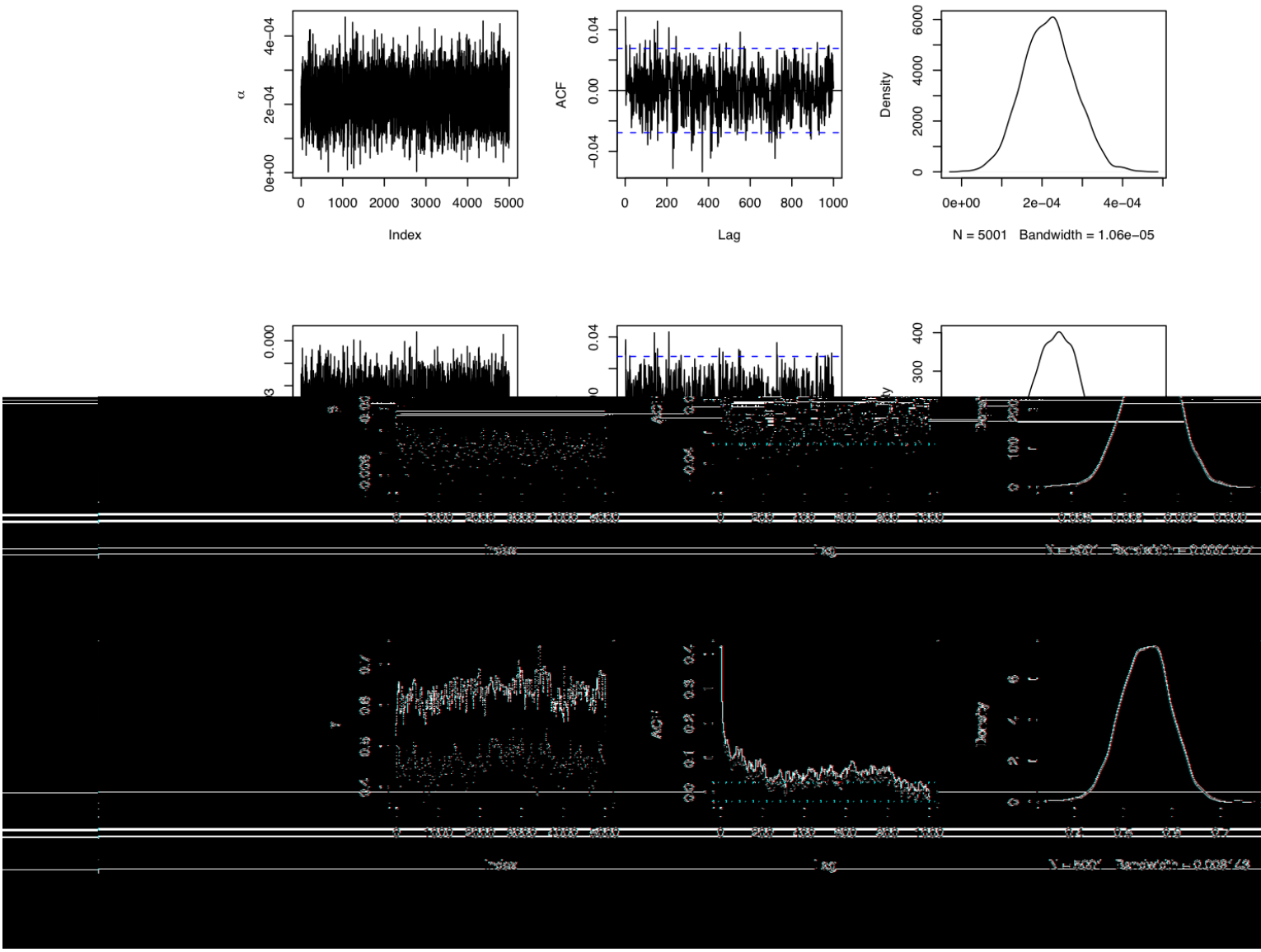

Figure 3.15: The simulation test of SVJ model–part 1: left column: the sample paths; middle column: the first 1000th autocorrelations (excluding the first); right column: posterior densities.

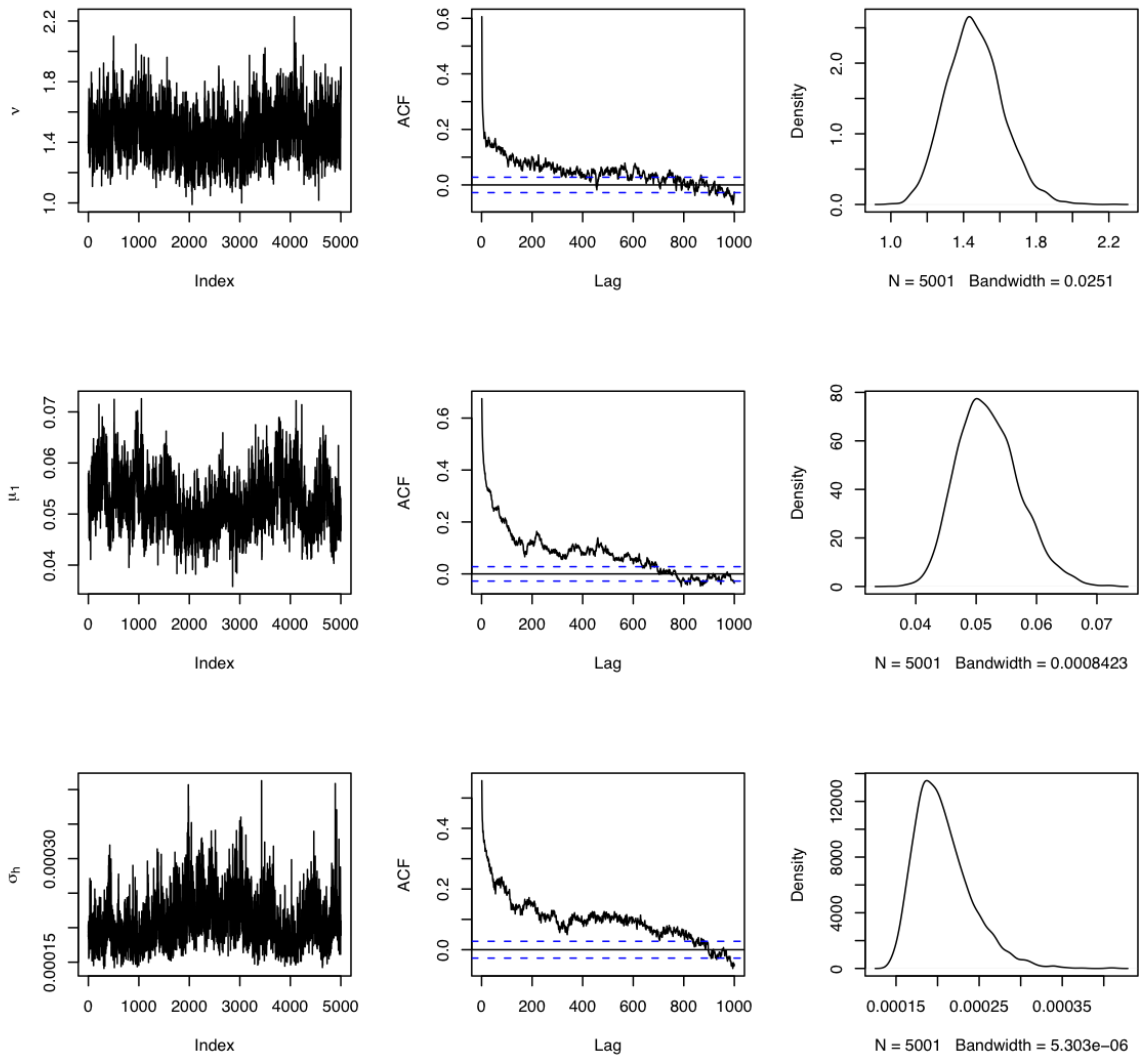

Figure 3.16: The simulation test of SVJ model–part 2: left column: the sample paths; middle column: the first 1000th autocorrelations (excluding the first); right column: posterior densities.

Table 3.5: Simulation Test for the SVS Model

Parameter	**True**	**Mean**	**Std.**	95% **C.I.**
α ($\times 10^{-4}$)	2.00	1.60	0.55	(0.52, 2.69)
β ($\times 10^{-3}$)	-3.00	-2.24	0.67	(-3.57, -0.94)
γ ($\times 10^{-1}$)	8.00	8.20	0.58	(7.08, 9.38)
σ_h ($\times 10^{-4}$)	2.00	1.99	1.85	(1.73, 2.44)
μ_1 ($\times 10^{-1}$)	7.00	5.68	1.04	(4.04, 8.06)
ν_1 ($\times 10^{-1}$)	1.50	2.04	0.48	(1.23, 3.13)
μ_2 ($\times 10^{-2}$)	3.00	3.06	0.32	(2.42, 3.70)
ν_2	1.00	0.85	0.15	(0.58, 1.20)

true values.

3.5.4 Simulation tests for the SVS model

The test results for the SVS model are listed in Table 3.5. The bias of γ is less significant than the SVJ model. For the parameters of latent process, both μ_1 and ν_1 are more biased against true values than the others. Since these are the parameters used to capture the short-term variation of latent volatility process which is an unstable process, these are acceptable results. All true values fall into corresponding 95% confidence interval of the SVS's parameters. In Figure 3.17, α and β are the fastest ones to converge to their stationary distribution and autocorrelations of their samples drop around zero just after a few lags. The other two quicker converging parameters are ν_1 and ν_2 observed from Figure 3.18.

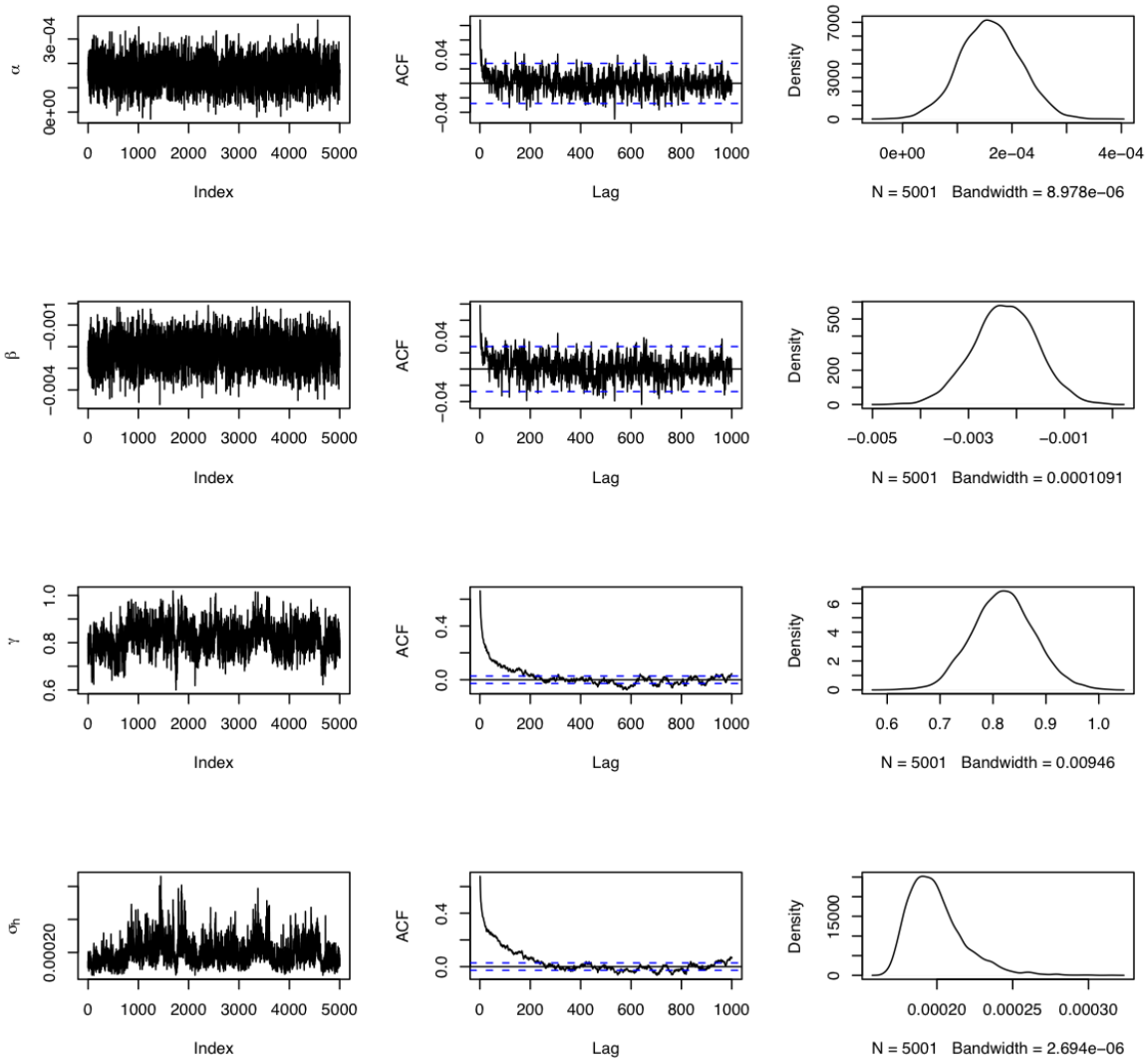

Figure 3.17: The simulation test of SVS model–part 1: left column: the sample paths; middle column: the first 1000th autocorrelations (excluding the first); right column: posterior densities.

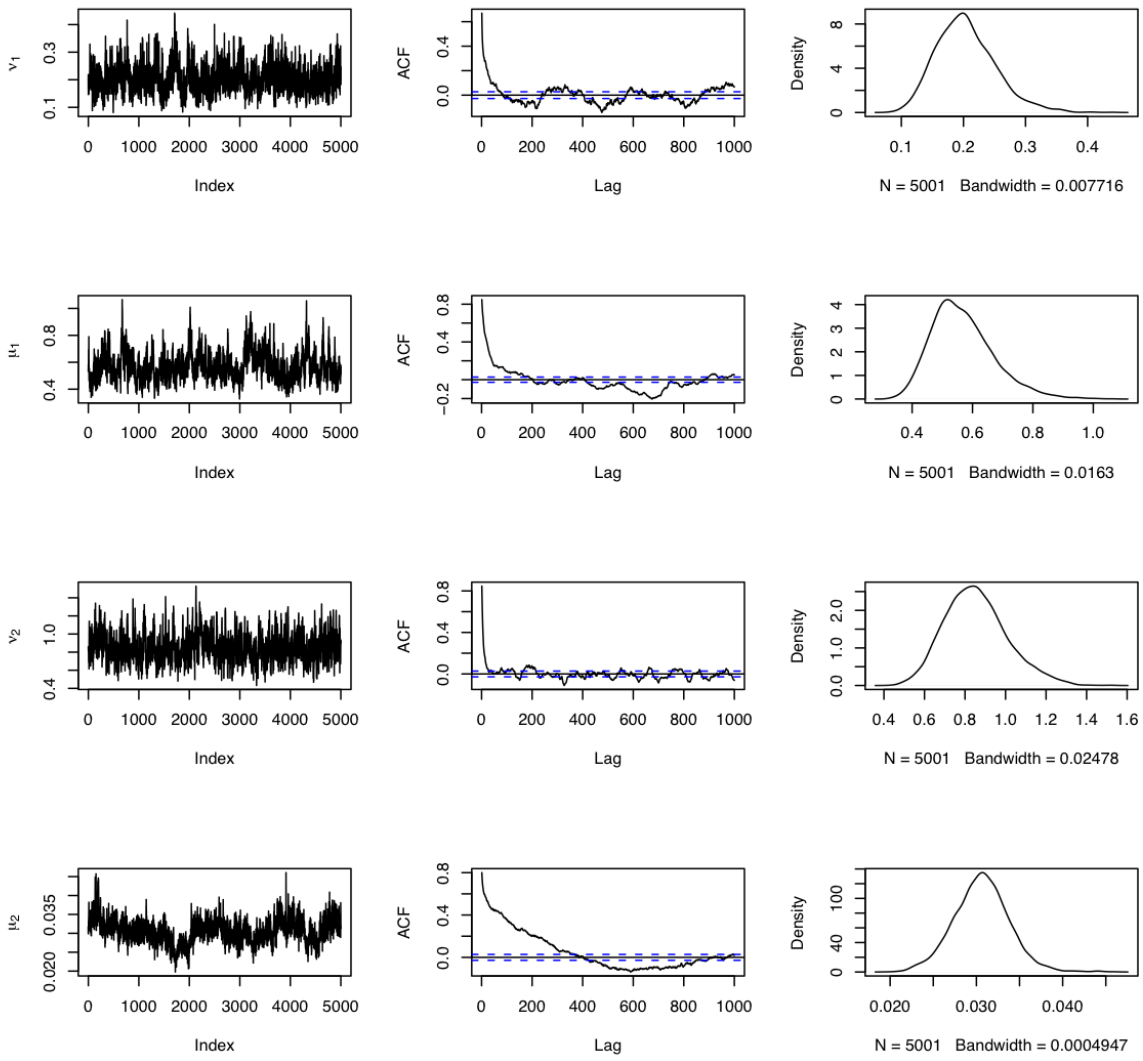

Figure 3.18: The simulation test of SVS model–part 2: left column: the sample paths; middle column: the first 1000th autocorrelations (excluding the first); right column: posterior densities.

Chapter 4

Empirical Study

4.1 Introduction

In this chapter we will present estimation results using the CKLS model, the SVG model, the SVJ and SVS model from the algorithms we have described in Chapter 3. For each parameter of every model, total 6 million MCMC iterations have been run. The first 2 million iterations are ignored as *burn-in* which are long enough to eliminate the effect of the starting points of each MCMC sample path. In order to decrease the dependency of simulation, the data we finally used to estimate are the thinned sample obtained by retaining every 100th iteration. The computer language for the implementation of the MCMC algorithm is C++. As we will see in following sections, the statistical diagnostics provide strong evidence for the superposition jumps in volatility.

4.2 Historical Data

The data which we try to fit with these models are the weekly rates of 3-month U.S Treasury Bill. These rates are constructed by picking up every Wednesday data from Federal Reserves H.15 reports of daily market rate: Wednesday having the fewest missing observations (Andersen & Lund (1997)). This sample then consists of 2242 weekly observations of yields in total, covering the period from January 1954

Table 4.6: Descriptive Statistics for 3-month T-Bill Weekly Yield

	Mean	**Std.**	**Skewness**	**Kurtosis**
weekly	*n=2242*			
r_i	0.055632	0.02866	1.084668	4.58413
$\triangle r_i$	0.000016	0.00255	-0.696971	23.68923
ϵ_i	0.001449	0.04097	0.135458	12.91045

to October 1997 including the most volatile period of interest rate: 1978-1982. As we will see, such a long period of observation offers us not only sufficient evidence of heteroscedasticity but also more information about unusual events, represented by jumps in our model. This sample has been previously analyzed in many academic studies using different models and estimation methods, see Andersen & Lund (1997) , and Eraker (2001) for example. So it is convenient for us to make comparison among these results using the same data.

Figure 4.19 presents the weekly evolutions of the 3-month T-Bill yield series $\{r_i\}, i = 0, ..., 2241$, its first differences (weekly change) $\{\Delta r_i\}$ where $\Delta r_i = r_{i+1} - r_i$, and the proportional difference (weekly change divided by the previous value), $\{\epsilon_i\}$ where $\epsilon_i = \Delta r_i / r_i$. There are three significant period s of volatility: 1954-1961, 1969-1975, and 1978-1972. In the first period, the dramatical fall in 1958 results in a significant change in the standardized difference series. The short rate reached its highest level in the third period which is the result of suddenly increase in oil prices in 1978.

The statistical properties of these observations are presented in Table 4.6. All of $\{r_i\}$, $\{\Delta r_i\}$ and $\{\epsilon_i\}$ have a high kurtosis, which indicates their distribution are fat tailed. It is also a evidence that the single-factor model is not a good candidate to fit the historical data. More stylized empirical facts for interest rate observed from this sample are presented below:

1. Non-negative: the evolution of interest rate never went below zero. In fact, high inflation might have caused the real rate to become negative. However,

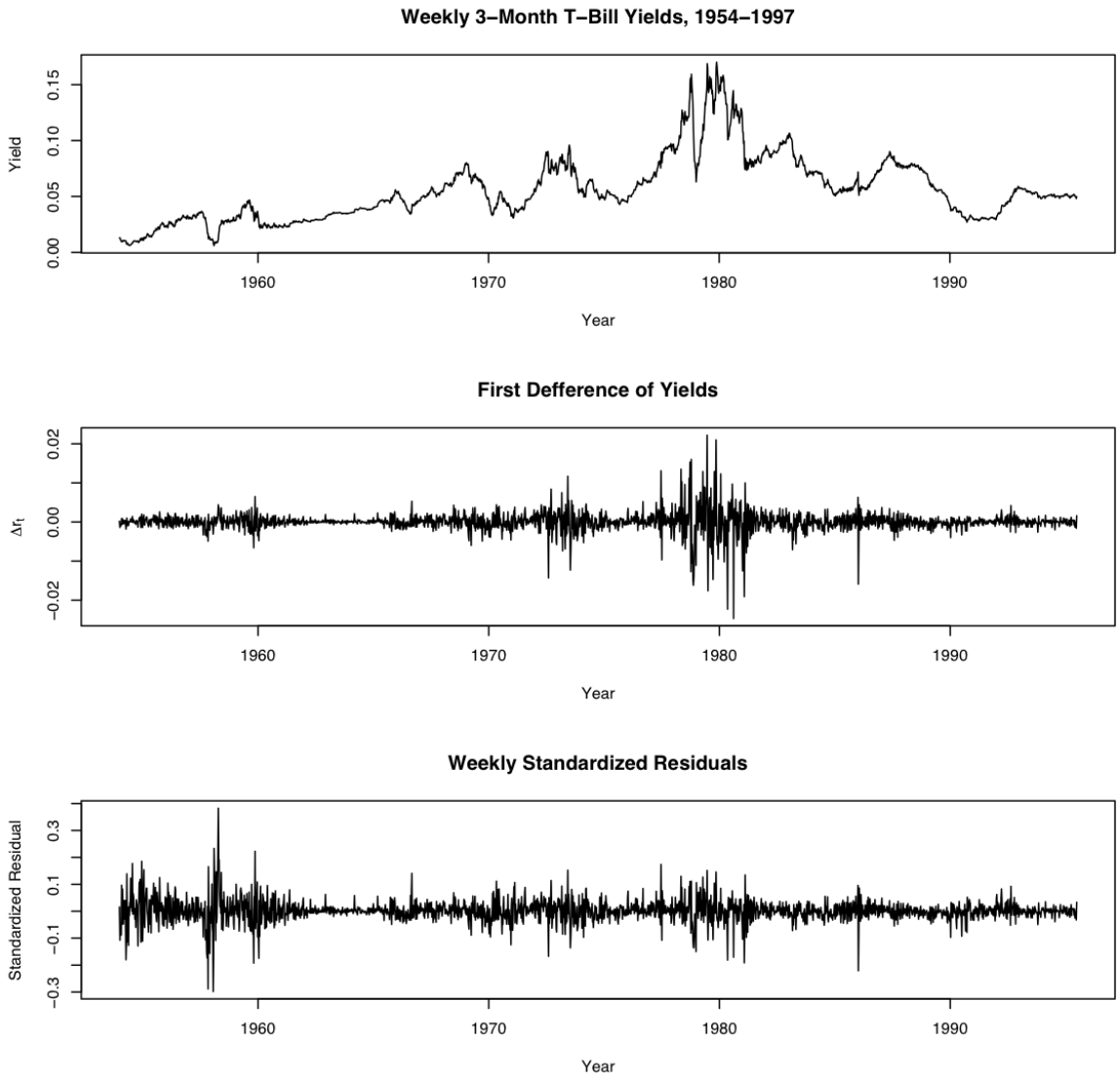

Figure 4.19: Historical Data. Top: Weekly 3-month T-bill yields covering the period from 1954 to 1997, selected from daily data; Middle: the first difference of interest rate, $\Delta r_i = r_{i+1} - r_i$; Bottom: Standardized residual, $\epsilon_i = \frac{\Delta r_i}{r_i}$ which shows strong clustering behaviour

we study the stochastic behaviour of short rate and ignore inflation.

2. Mean-reversion: which means interest rates appear to be pulled back to some long-term average level over time. The lower short rate at 1955 had been pulled up to a higher level in the following years; and the highest rate around 80s' moved to lower level.

3. Heteroscedasticity (Volatility clustering): the phenomenon of 'periods of high volatility followed by periods of low volatility' is very significant in both the middle and bottom graphs of Fig 4.19.

4. Level effect: the corresponding volatility for period 1978-1982 increased obviously as the short rate attained the highest level.

5. Excess kurtosis and skewness: the results from Table 4.6. indicate that changes in the short rate have a fatter tail density than the Normal distribution.

These stylized empirical facts can be benchmarks for the model selection. A desirable model should be capable to capture these characteristics observed in the market. Recall the short rate models we described in Chapter 1 and Chapter 2, the drift part of the models is designed to capture the mean-reversion characteristics. The non-negative property of interest rate can be obtained through the component, r_t^γ, by setting $\gamma = 0.5$ or $\gamma = 1$ for example. Chan *et al.* (1992) attempt to explain the heteroscedasticity by making the volatility proportional to r_i^γ. However, as we will show in the sections that follow, this is not sufficient to explain the pattern of stochastic volatility in the data. Both volatility clustering and excess kurtosis can be achieved by introducing the stochastic volatility process. Improvements can be obtained by changing the diffusion-based SV model to the model incorporating with jumps.

Table 4.7: Posterior parameter estimates for the CKLS model

Parameter	**Mean**	**Std.**	**t-stat**
α	1.601 $\times 10^{-4}$	0.557 $\times 10^{-4}$	2.875
β	-2.512 $\times 10^{-3}$	1.413 $\times 10^{-3}$	-1.776
γ	0.763	0.018	41.991
σ_r	1.931 $\times 10^{-2}$	0.109 $\times 10^{-2}$	17.647
Marginal Log-Likelihood			10851.8

4.3 Estimation Results for the CKLS Model

4.3.1 Parameter estimations

Recall the discrete-time version of the CKLS model is

$$\Delta r_i = (\alpha + \beta r_i)\Delta t + \sigma_r r_i^{\gamma} \Delta W_i,$$

where $\Delta r_i = r_{i+1} - r_i$ and ΔW_i is a Brownian motion with variance equal to Δt which is 1 week in our case. Table 4.7 presents the estimation results of the CKLS model using the MCMC method. Alongside the posterior mean of each parameter, both standard deviation and t-statistics are also listed. All the parameters are significantly unlike zero. Note also that $\gamma = 0.763$ is much smaller than the result of Chan *et al.* (1992) (around 1.5 from monthly data). The mean reversion property is not rejected by CKLS model with no-zero β. The long term mean of 3-month T-bill yield calculated from the CKLS model, $-\alpha/\beta$ is about 6.37%. The corresponding annualised estimations are $\alpha = 8.33 \times 10^{-3}$, $\beta = -0.833$, $\sigma_r = 0.14$ and $\gamma = 0.763$.

4.3.2 Residual analysis

By the assumption of the short rate differential equation, the CKLS residuals which are defined as:

$$\xi_i := \frac{\Delta r_i - (\alpha + \beta r_i)\Delta t}{\sigma_r r_i^{\gamma} \sqrt{\Delta t}}$$

should be *i.i.d.*, following the standard Normal distribution. We will test both the normality and independency of the residuals to assess the fitting ability of the CKLS

Table 4.8: Moments of residuals, ξ and the JB statistics

	CKLS	**SVG**	**SVJ**	**SVS**
Mean	0.00060	0.00769	0.02400	0.01555
Std.	0.99756	0.87781	1.27134	1.04941
Skewness	-0.22297	-0.07485	0.17035	0.00829
Kurtosis	9.69371	4.45598	3.76705	3.19300
JB-test	4202.31	229.273	65.7779	3.50379

model. The parameters used are the estimation results listed in Table 4.7.

Normality of residuals

The top plot in Figure 4.20 presents the residual process of the CKLS model. This is obviously not a i.i.d. process. Thus, even incorporating volatility with interest rate level, r^γ, CKLS is not flexible enough to capture the heteroscedasticity. The observation of Figure 4.19 shows that one of the high volatility period 1950-1960 does not accompany with high levels of interest rate. Both indicates the level effect alone cannot explain the randomness of interest rate volatility.

Further normality diagnostics of the residuals are shown in the Quantile-Quantile (Q-Q) plot and the sample densities. In the middle plot of Figure 4.20, the inverted S shape of Q-Q plot is caused by concentrated small values and large outliers which indicates that the residuals of CKLS have a fatter tail distribution as shown in bottom plot of Figure 4.20. Both Q-Q plot and the graph of comparison between empirical density of the residuals against standard normal distribution for the CKLS model suggest to reject the null hypothesis of normality.

First column of Table 4.8 summarizes the means, standard deviations, skewness and kurtosis of the CKLS residuals . Under the normal hypothesis, the values of the moments of residuals should be $0, 1, 0$ and 3 respectively. Whereas, CKLS has negative skewness which suggests a slightly left-skewed density and leptokurtosis.

With these information, we will apply the Jarque-Bera (JB) test to diagnose the null hypothesis of normality. The JB test has a goodness-of-fit measure of departure from normality based on the sample skewness and kurtosis. The JB statistic is defined as following:

$$JB = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right)$$

where n is the sample size; s presents skewness and k presents kurtosis. Then JB has a chi-squared distribution with 2 degrees freedom under the null hypothesis, $H_0 : s = 0, k = 3$. The results are provided under the bottom of Table 4.8. As can be seen, even with 99.5% confident interval, the CKLS model fails the JB test.

Independence of residuals

The first 100 autocorrelations (except the first one) of the CKLS residuals are presented in the top plot of Figure 4.21. Around twenty percent autocorrelations are out of the 95% confidence interval (the dot line). The pattern of autocorrelations for the squared residuals strongly reject the independence assumption shown in the middle plot of Figure 4.21. The inability of CKLS model to fit historical data is due to the simple structure of both drift and volatility part. Nonlinear drift structure has been studied in many literature, such as Aït-Sahalia (1996) and Andersen & Lund (1997). The inclusion of a more complex volatility structure will be discussed in the following sections.

The bottom plot of Figure 4.21 presents further diagnostic test to investigate the i.i.d. hypothesis of residuals. As we know if a time series is i.i.d., then the times between exceedances of a given threshold should have a geometric distribution. When we choose the 95% quantile of each series as the threshold respectively, and calculate the waiting times between exceedances then these should come from a Geometric distribution with probability $p = 0.05$. As seen of the bottom plot in Figure 4.21, the short term waiting times are smaller than the values predicted by the geometric distribution and the long term ones are above the straight line. One reason for the

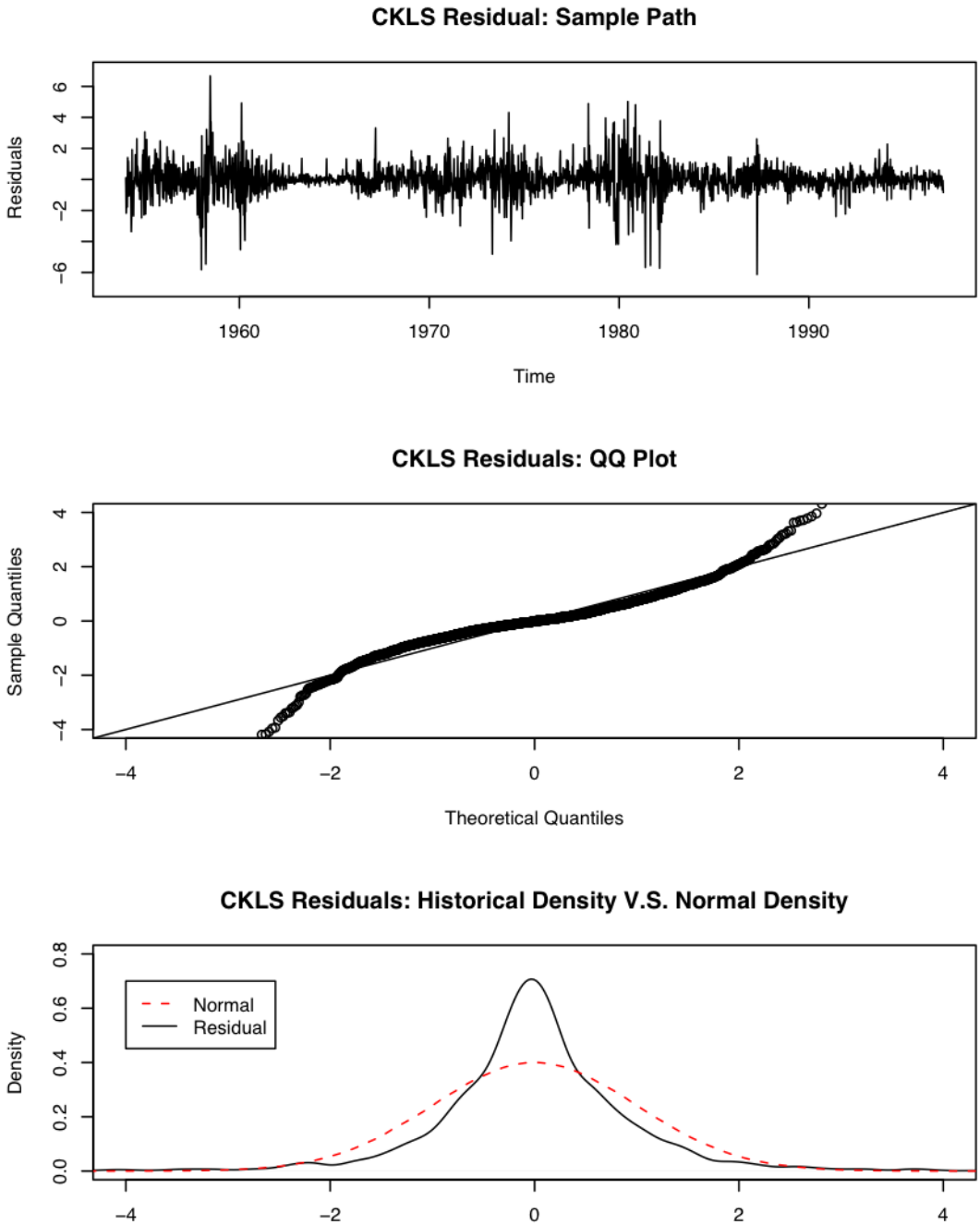

Figure 4.20: Tests of Normality for the CKLS Residual

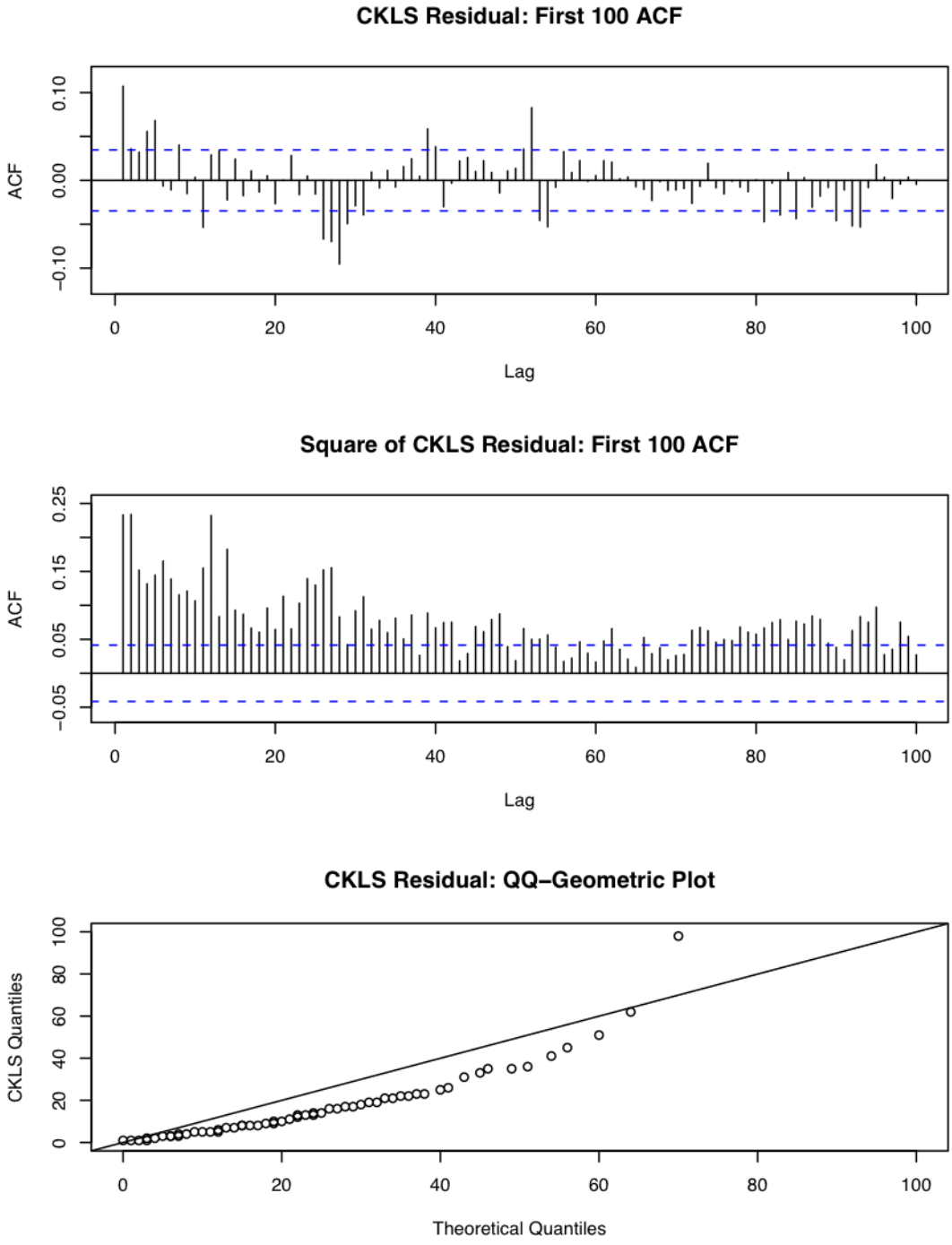

Figure 4.21: Test of I.I.D. for the CKLS Residual

Table 4.9: Posterior parameter estimates for the SVG model

Parameter	**Mean**	**Std.**	**t-stat**
α	1.111 $\times 10^{-4}$	0.552 $\times 10^{-4}$	2.011
β	-1.754 $\times 10^{-3}$	0.833 $\times 10^{-3}$	-2.106
γ	0.714	0.164	4.337
σ_r	1.367 $\times 10^{-2}$	0.681 $\times 10^{-2}$	2.007
σ_h	0.436	0.053	8.230
μ	0.053	0.013	4.082
Marginal Log-Likelihood			11432.9

S-shape of the Q-Q plot is due to the clusters of volatility in CKLS residuals. For the stochastic volatility models, shown in other plots in the same figure, most of the times adhere to the predicted ones with few outliers in the long time area. Therefore, the Q-Q Geometric plot indicates that the calculations from the SV models give a better explanation of the historical data than the CKLS model.

4.4 Estimation Results for the SVG Model

4.4.1 Parameter estimations

Recall the discrete-time version of the SVG model described in section 3.3 is

$$\begin{aligned} \Delta r_i &= (\alpha + \beta r_i)\Delta t + \sigma_r e^{\frac{1}{2}\hat{h}_i} r_i^{\gamma} \Delta W_{1,i}, \\ \Delta \hat{h}_i &= -\mu \hat{h}_i \Delta t + \sigma_{\hat{h}} \Delta W_{2,i}, \end{aligned}$$

where $\hat{h}_i$ is defined in section 2.4, $\Delta W_{1,i}$ and $\Delta W_{2,i}$ are two independent Brownian motions with zero mean and variance Δt which is equal to 1 (week) in our case. Table 4.9 presents the estimation results of the SVG model. The mean reversion property is not rejected by the SVG model either, where β is significant unlike zero. The long term mean, $-\alpha/\beta$ from the SVG model decreases slightly from 6.47% to 6.33%. The value of γ declines to 0.714 comparing with 0.763, the one of the CKLS model. The small converge rate μ indicates the high persistence in stochastic

volatility process. Since its value is also significantly unlike zero, the mean reversion property has been kept in the volatility process. The corresponding annulised estimations are $\alpha = 5.78 \times 10^{-3}$, $\beta = -9.12 \times 10^{-2}$, $\sigma_r = 9.86 \times 10^{-2}$, $\sigma_h = 3.14$, $\mu = 2.76$ and $\gamma = 0.714$.

4.4.2 Residual analysis

The SVG residual is defined as:

$$\xi_i := \frac{\Delta r_i - (\alpha + \beta r_i)\Delta t}{\sigma_r e^{1/2h_i} r_i^{\gamma} \sqrt{\Delta t}},$$

for $i = 0, ..., 2241$. The parameters used here are the results listed in Table 4.9 and $\{h_i\}$ is one of the stationary process in the MCMC iteration. According to the model assumption, $\{\xi_i\}$ should be *i.i.d.*, following the standard Normal distribution, and we will test both the normality and independence of the residuals to assess the fitting ability of the SVG model.

Normality of residuals

Figure 4.22 presents normality diagnostic tests of the SVG residuals. The improvement of the fitness of SVG model is significant. The SVG residuals move more similarly as Wiener process than CKLS model as seen from the top plot. However, there is a slight inverted S-shape in the Q-Q plot and fat tails residual density provide evidence about the inability of Gaussian driven random process in capturing volatile behaviour of short rate volatility. The same conclusion can be obtained by examining the statistics in Table 4.8. The kurtosis of SVG residual is still bigger than idea value of 3 although it is smaller than CKLS's. From the bottom of second column, the SVG model also fails the JB test.

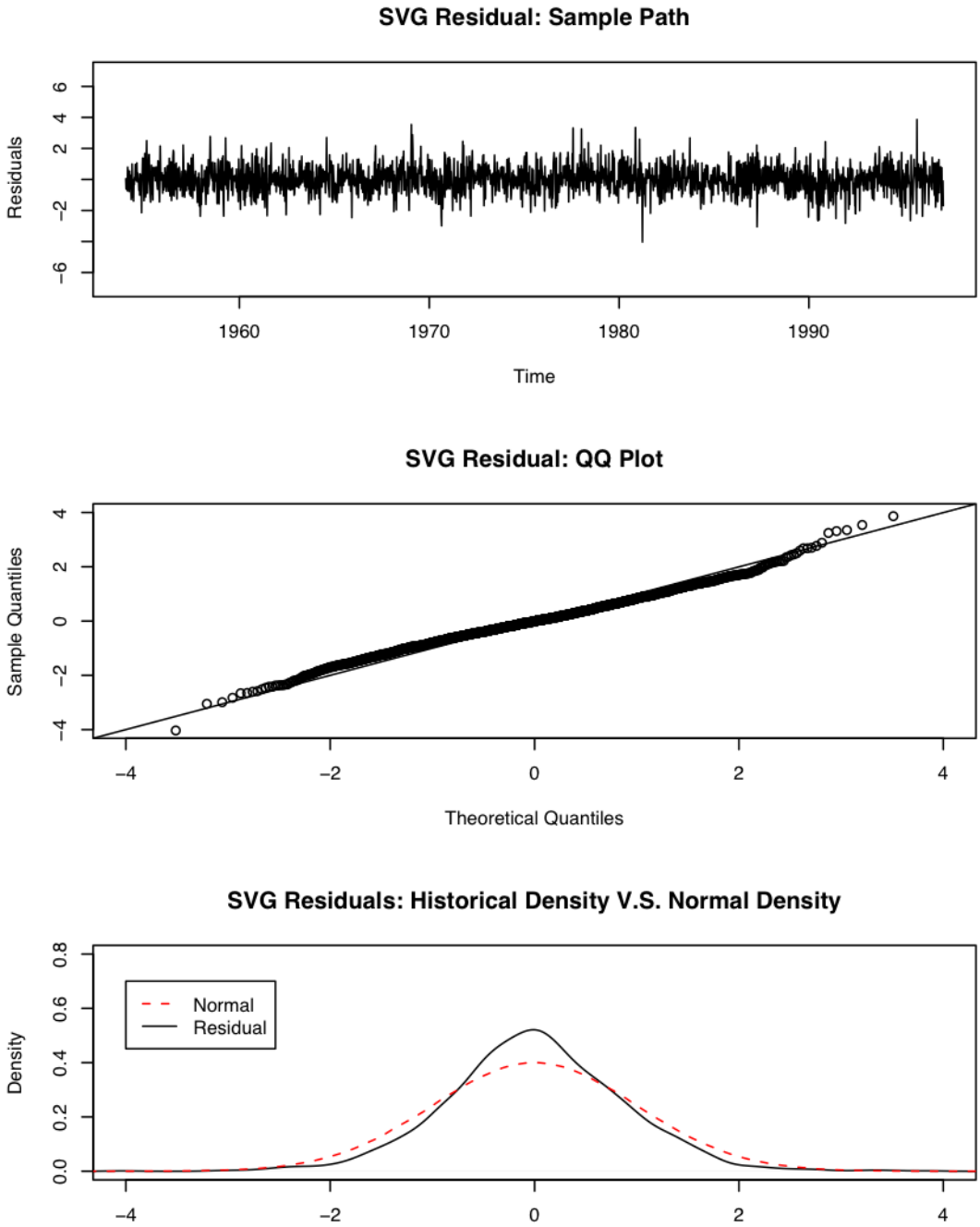

Figure 4.22: Tests of Normality for the SVG Residuals

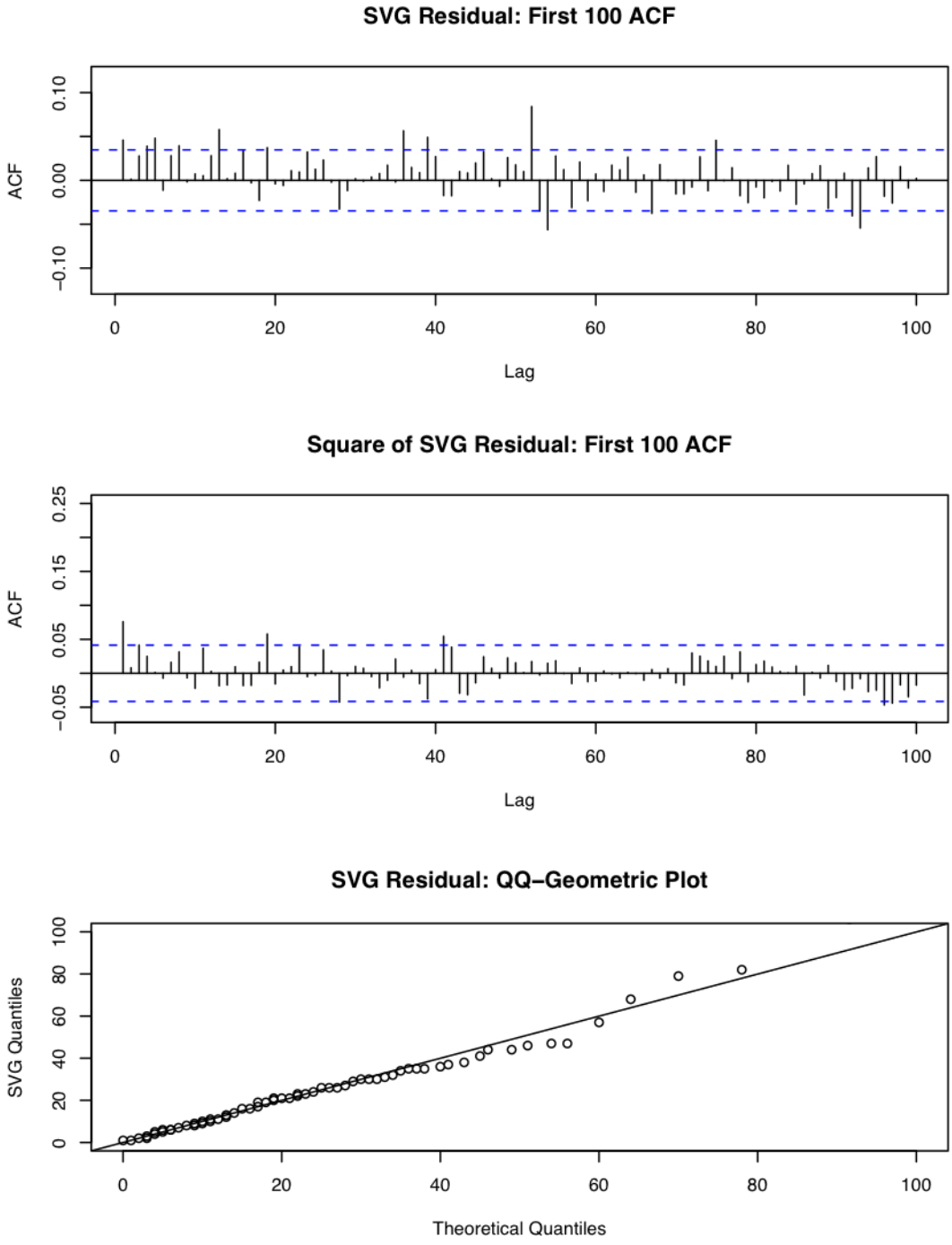

Figure 4.23: Tests of I.I.D. for the SVG Residuals

Table 4.10: Posterior parameter estimates for the SVJ model

Parameter	**Mean**	**Std.**	**t-stat**
α	1.393 $\times 10^{-4}$	0.491 $\times 10^{-4}$	2.834
β	-2.379 $\times 10^{-3}$	1.124 $\times 10^{-3}$	-2.117
γ	0.642	0.069	9.222
σ_h	1.437 $\times 10^{-4}$	0.654 $\times 10^{-4}$	2.197
μ	0.041	0.005	8.823
ν	1.283	0.188	6.832
Marginal Log-Likelihood			11485.3

Independence of residuals

Comparing to the CKLS model graphical diagnostic tests for the independence of residuals are also better for the SVG model. In the top plot of Figure 4.22, there are only 11% points over the the 95% confidence interval in the first 100 autocorrelations. The amplitude of these outliers is much less than those of CKLS model. This improvement is more clearly revealed by the ACF of squared residuals. These autocorrelations are small and fluctuating around zero. The Q-Q Geometric plot in bottom of Figure 4.23 gives another evidence for the relative suitability of SVG model. The points beyond 40 are fluctuating around the line, and the short-term waiting times (points before 40) between the 95% threshold are very close to the values from the Geometric distribution.

4.5 Estimation Results for the SVJ Model

4.5.1 Parameter estimations

Recall in section 3.4, the discrete form of the SVJ model is written as:

$$\begin{aligned}
\Delta r_i &= (\alpha + \beta r_i)\Delta t + \sqrt{v_i(\Delta t)} r_i^{\gamma} \Delta W_i, \\
v_i(\Delta t) &= \frac{1}{\mu}\left(h_i \left(1 - e^{-\mu \Delta t}\right) + \sum_{j=1}^{N_i(\Delta t)} \varepsilon_j \left(1 - e^{-\mu((i+1)\Delta t - c_j)}\right)\right), \\
h_i &= h_0 e^{-\mu i \Delta t} + \sigma_h \sum_{j=1}^{N(i\Delta t)} e^{-\mu(i\Delta t - c_j)} \varepsilon_j
\end{aligned}$$

where ΔW_i is a Brownian motion with variance Δt (1 week here), $v_i(\Delta t)$ is the integrated volatility, $N_i(\Delta t) = N((i+1)\Delta t) - N(i\Delta t)$ is the number of jumps from the ith week to the $i+1$th week, and $N_0(i\Delta t) \equiv N(i\Delta t)$. (c_j, ε_j) are the corresponding jump points and the jump size $\varepsilon_j \sim Exponential(\nu)$.

The estimation results of the SVJ model are listed in Table 4.10. As indicated from the *t-statistics*, all parameters are significantly unlike zero. The value of short rate converge rate, β, is between the CKLS and SVG model. The long term mean, however, is 5.855%, the smallest one in the results we have given so far. The value of γ drops to 0.642, smaller than the other two. The jump volatility model is also a highly persistent process as the SVG model with the small value of μ. Given the posterior values of μ and ν, the value of weekly jump rate $\lambda = \mu\nu$ is 0.052603 which indicates that we should expect around 3 jumps each year. The corresponding annualised estimations are $\alpha = 7.24 \times 10^{-3}$, $\beta = -0.124$, $\sigma_h = 2.761 \times 10^{-6}$, $\mu = 2.132$, $\nu = 1.283$ and $\gamma = 0.642$.

4.5.2 Residual analysis

The SVJ residuals are defined as:

$$\xi_i := \frac{\Delta r_i - (\alpha + \beta r_i)\Delta t}{\sigma_r \sqrt{v_i(\Delta t)} r_i^{\gamma} \sqrt{\Delta t}}.$$

Parameters used are the results listed in Table 4.10 and the integrated volatility v_i is one stationary process from the MCMC iterations. Following the assumption, $\{\xi_i\}$ should be *i.i.d.*, following the standard Normal distribution. We will test both the normality and independence of the residuals to assess the fitting ability of the SVJ model.

Normality of residuals

Figure 4.24 provides the evolution plot of SVJ residuals, the Q-Q plot and the density function. As we can see from the top plot, although the values of SVJ residuals aggregate mostly in the interval of $(-2, 2)$, there are many more points reaching -4 or 4 than the SVG residuals. Therefore we will expect a fat-tailed density compared with standard normal distribution. Then the Q-Q plot is straighter than the S-shape for the previous two models. But compared with the standard normal density, the SVJ model does not provide a satisfying solution. From the third column of Table 4.8, an improvement resulting from the use of a different volatility process is not very clear. The kurtosis decreases to 3.77 but both the standard deviation and mean are further from the theoretical value. As a result, this model still cannot pass the JB test.

Independence of residuals

The first 100 autocorrelations of both SVJ residuals and squared residuals are plotted in the top and middle graphs in Figure 4.25. Similar to the SVG model, the position of residual ACF over 95% confidence interval is only around 10%. For the squared residuals, the first 100 autocorrelations are also around zero. Both the diffusion-based model and the jump model provide reasonable ACF patterns. In the Geometric test for the waiting time, the departure of the SVJ outliers only appears at the middle area which provide some evidence in favour of the jump process over the continuous Brownian motion. Further tests are given in later sections and the improvement of jump process will be more clear.

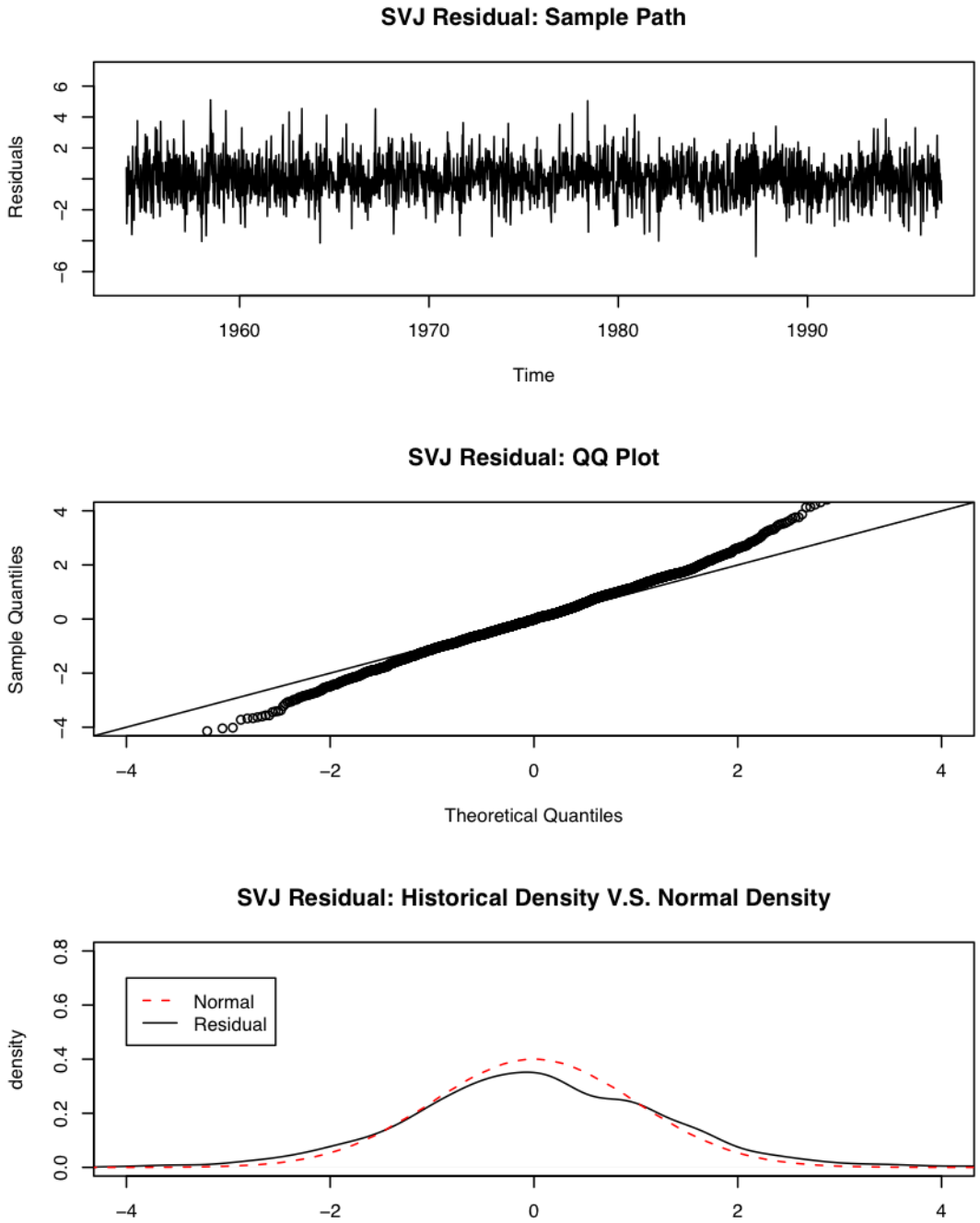

Figure 4.24: Tests of Normality for the SVJ Residuals

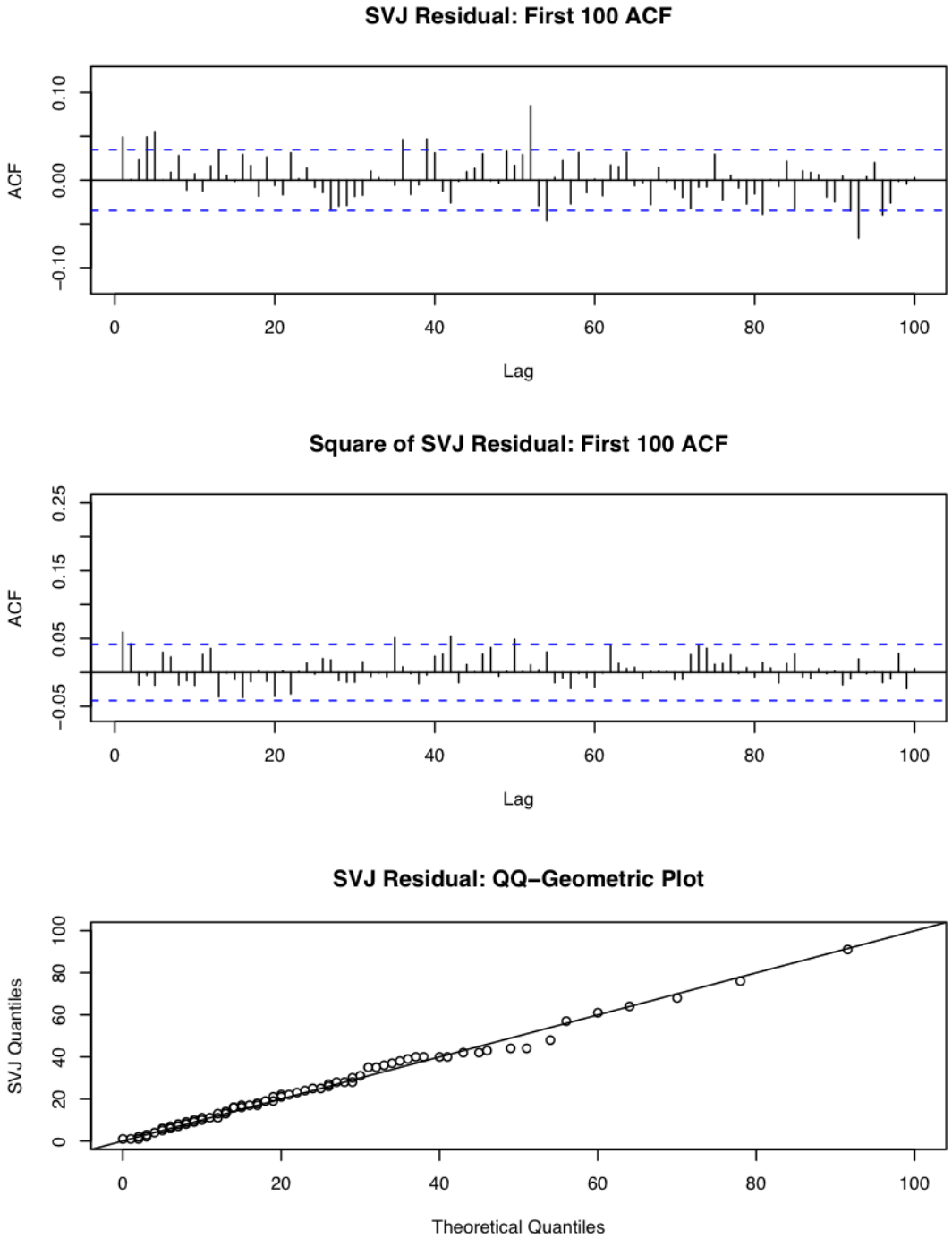

Figure 4.25: Tests of I.I.D. for the SVJ Residuals

Table 4.11: Posterior parameter estimates for the SVS model

Parameter	**Mean**	**Std.**	**t-stat**
α	1.096 $\times 10^{-4}$	0.497 $\times 10^{-4}$	2.204
β	-1.774 $\times 10^{-3}$	0.891 $\times 10^{-3}$	-1.991
γ	0.672	0.072	9.283
σ_h	2.067 $\times 10^{-4}$	1.028 $\times 10^{-4}$	2.107
μ_1	0.729	0.339	2.152
ν_1	0.150	0.053	2.848
μ_2	0.029	0.004	8.100
ν_2	0.979	0.164	5.972
Marginal Log-Likelihood			11695.2

4.6 Estimation Results for the SVS Model

4.6.1 Parameter estimations

Recall that the discrete form of the SVS model used in MCMC algorithm is:

$$\Delta r_i = (\alpha + \beta r_i)\Delta t + v_i(\Delta t)^{1/2} r_i^{\gamma} \Delta W_i,$$

where ΔW_i is a Brownian motion with variance Δt which is 1 (week) here, $v_i(\Delta t)$ is the integrated volatility defined as below:

$$v_i(\Delta t) = v_{1,i}(\Delta t)) + v_{2,i}(\Delta t),$$

$v_{m,i}(\Delta t)$ the integrated volatility for each part with $m = 1, 2$ is

$$v_{m,i}(\Delta t) = \frac{1}{\mu_m} \left(h_{m,i} \left(1 - e^{-\mu_m \Delta t}\right) + \sum_{j=1}^{N_{m,i}(\Delta t)} \varepsilon_{m,j} \left(1 - e^{-\mu_m((i+1)\Delta t - c_{m,j})}\right) \right),$$

and $h_{m,i}$ for $m = 1, 2$ is

$$h_{m,i} = h_{m,0} e^{-\mu_m i \Delta t} + \sigma_h \sum_{j=1}^{N_m(i\Delta t)} e^{-\mu_m(i\Delta t - c_{m,j})} \varepsilon_{m,j}.$$

$N_{m,i}(\Delta t) = N_m((i+1)\Delta t) - N_m(i\Delta t)$ is the number of jumps happening from the ith week to the $i+1$th week for the m part. $(c_{m,j}, \varepsilon_{m,j})$ are the corresponding jump

points and the jump size $\varepsilon_{m,j} \sim Exponential(\nu_m)$.

The estimation results of the SVS model are listed in Table 4.11. It is worth to notice that most of the parameter values are lying between the ones of SVG and of SVJ model. The converging parameters β, μ_1 and μ_2 are all significant from zero which indicate that the mean reversion property is not rejected in any part. The long term mean from this model is 6.178%. Another parameter lying between the other two SV models is γ and its value is also far below the results of Chan *et al.* (1992). For the jump process with faster converge rate, the weekly jump rate $\lambda_1 = \mu_1\nu_1$ is 0.109 indicating an average of almost 6 jumps per year. For the higher persistence part, there are about 1 jump per year since the weekly jump rate for this part $\lambda_2 = \mu_2\nu_2$ is only 0.0284. Obviously, small jumps dominate the volatility process. As with previous sections, we will assess the SVS model by testing the residuals. The corresponding annualised estimations are $\alpha = 5.699 \times 10^{-3}$, $\beta = -9.22 \times 10^{-2}$, $\sigma_h = 3.975 \times 10^{-2}$, $\mu_1 = 37.9$, $\mu_2 = 1.508$, $\nu_1 = 0.15$, $\nu_2 = 0.979$ and $\gamma = 0.672$.

4.6.2 Residual analysis

The SVS residuals are defined quite similar as the SVJ model:

$$\xi_i := \frac{\Delta r_i - (\alpha + \beta r_i)\Delta t}{\sigma_r \sqrt{v_i(\Delta t)} r_t^{\gamma} \sqrt{\Delta t}}.$$

The parameters used in calculation are the estimation results listed in Table 4.11, The integrated volatility $v_i(\Delta t)$ is calculated given one stationary volatility process in the MCMC iteration. Under the model assumption, $\{\xi_i\}$ should be *i.i.d.* random variables from the standard Normal distribution. We will test the normality and independency of the residuals to assess the fitting ability of the SVS model.

Normality of residuals

Improvements resulting from the use of the SVS model are quite apparent from Figure 4.26. Firstly, the sample path in the top graph looks more similar as the

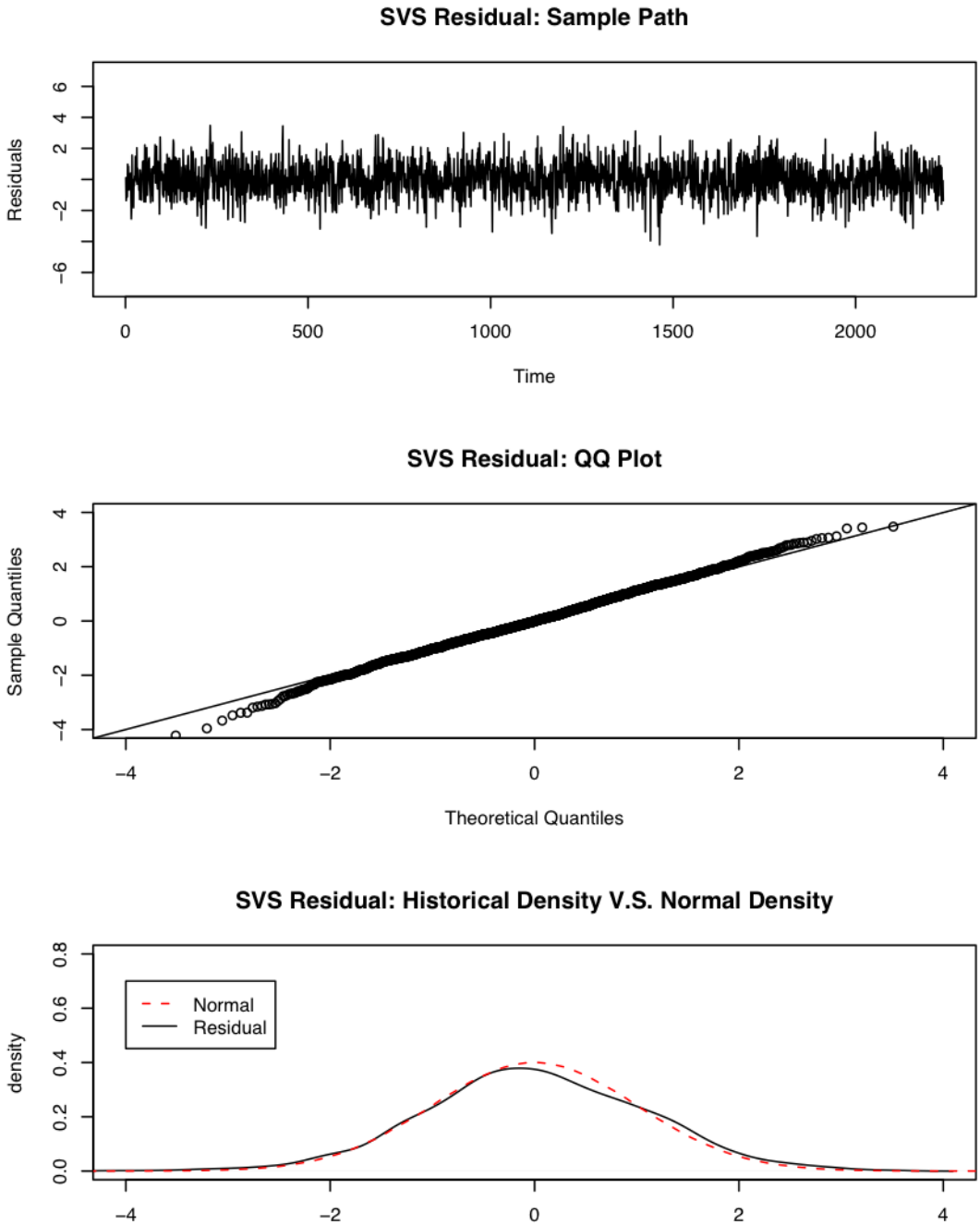

Figure 4.26: Tests of Normality for the SVS Residuals

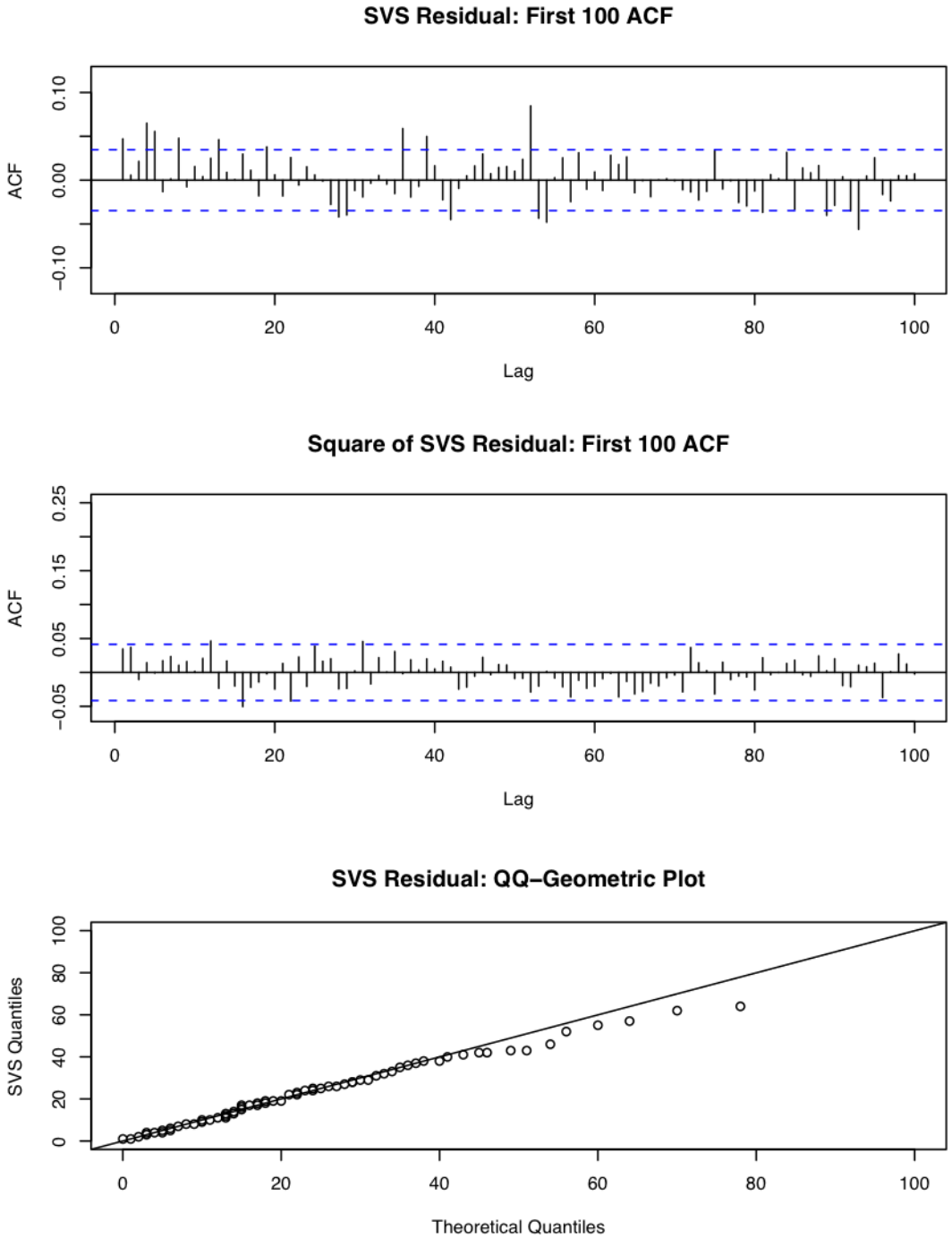

Figure 4.27: Tests of I.I.D. for the SVS Residuals

process of Gaussian increments than the others, where the SVS residuals fluctuate more widely than the SVG model while do not visit to the "large" values as the SVJ residuals do. Secondly, the Q-Q points of the SVS residuals are closer to the 45-degree line except some steepening at the left end. From the bottom graph, the comparison of densities give us a clear view of how close the residuals are to the Normal increment. Finally, as we examine the last column of Table 4.8, all the key moments are very close to the true values. The value of JB statistic is only 3.50, which is small enough to make the SVS model be the only candidate passing the test.

Independence of residuals

The *i.i.d.* tests for the SVS residuals are presented in Figure 4.27. As the other two SV models, the introduce of stochastic volatility process decreases the dependency among the residuals. The first 100 autocorrelations of both residuals and squares are small and fluctuate around zero. From the Q-Q Geometric plot at the bottom, the majority points which value are less than 40 are sticking to the straight line while few higher value points are just below the line. We have also observed similar patterns from the other two SV models.

Although the autocorrelation plots and the Q-Q Geometric test reveal the misspecification of CKLS model, neither can reveal which volatility model performs better. One of the well known facts is that uncorrelated random variables are not necessarily independent. That explains the rejections of both SVG and SVJ model in Chi-Square test appearing in a later section. The shortcoming of Q-Q Geometric test is the waste of data resource. The lack of filtered data presents blurred shapes around the short time period and dispersive points for the long waiting time area. Both of them are therefore not sensitive tests to investigate the SV models. A more sophisticated *i.i.d.* test will be given in the next section.

4.7 Copula

4.7.1 The basis

A *copula* $C(x)$ for an $n \times 1$ random vector X is a multivariate cumulative distribution function where the marginal distribution for the ith element x_i is a uniform distribution on $[0, 1]$, $i = 1, ..., n$. The copula is introduced to analyze the dependence among the random series. Different with the method we have applied in previous sections, copula will focus on the dependence structure and remove the marginal effect of random variables. Therefore, copula will offer us a deep view of dependence. In our purpose, we will only consider the bivariate case where $n = 2$. More discussions and properties of copula can be found in A.J.McNeil *et al.* (2005). The following theorem tell us the relationship among the joint distribution, the copula and the margins.

Theorem 4.1 ***(Sklar,1959)*** *Let F be a bivariate joint distribution function with margins F_1, F_2. Then there exists a copula $C : [0, 1]^2 \to [0, 1]$ such that, for all x_1, x_2,*

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \tag{4.82}$$

Moreover, if the margins are continuous, the copula function C is unique and for all $(u_1, u_2)' \in [0, 1]^2$,

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)). \tag{4.83}$$

Conversely, if C is a copula and F_1, F_2 are univariate distributions then the function F defined in 4.82 is a joint distribution with margins F_1, F_2.

Thus a copula can be constructed by the joint distribution and margins using the Sklar's theorem. Meanwhile, the joint distribution can also be found given the copula and margins of the random vectors.

4.7.2 Copulas of the residuals

For each candidate model, the corresponding residual $\{\xi_t\}$ should be *i.i.d.* random variables and their marginal distributions are the standard Normal distribution, Φ. The copula of $(\xi_t, \xi_{t+k})'$ should therefore be a sample from the *independence copula*. Let $U(t) = \Phi(\xi_t)$ be the probability transformed variables. The residual copula should have the structure as below:

$$\begin{aligned} C(\Phi(\xi_t), \Phi(\xi_{t+k})) &= \Phi(\xi_t)\Phi(\xi_{t+k}) \\ &= U(t)U(t+k). \end{aligned}$$

We refer the copula $C(U(t), U(t+k))$ as the kth copula of ξ_t for simplicity. For every non-zero k, we expect a uniformly distributed scatter plot of $(u(t), u(t+k))$ if ξ_t the residuals are *i.i.d.* Normal variables.

Figure 4.28 presents the scatter plots of the first copula for each model. For the CKLS model, the top-left plot shows that there are around 37% points gathered in the center area, $(0.3, 0.7) \times (0.3, 0.7)$, more than two times of the corresponding ideal value 16%. The frequencies of the points visit both the top-right and down-left corners are obviously higher than the other two corners. The copula plot for the SVG model is given in the top-right graph. Although the position of central gathering points drop from 37% to 25%, it is still bigger than the ideal percent. Another shortcoming of the SVG model is that there are too few points in the corners. The bottom two graphs are the copula plots for the SVJ model and the SVS model respectively. The benefit of changing the volatility process can now be seen clearly. The plots for both jump models are spread more uniformly than the previous two models. However, the points of the SVJ model visit borders quite often and there are few points left in the central area, 12%, seen from the down-left picture. Whereas, the plot for the SVS model is the one mostly closing to the uniform distribution, and there are 15% points falling in the central area.

Let us examine long distance copulas for these models, for example the 500th bivariate copulas presented in Figure 4.29. As the time-lag increases, the dependency

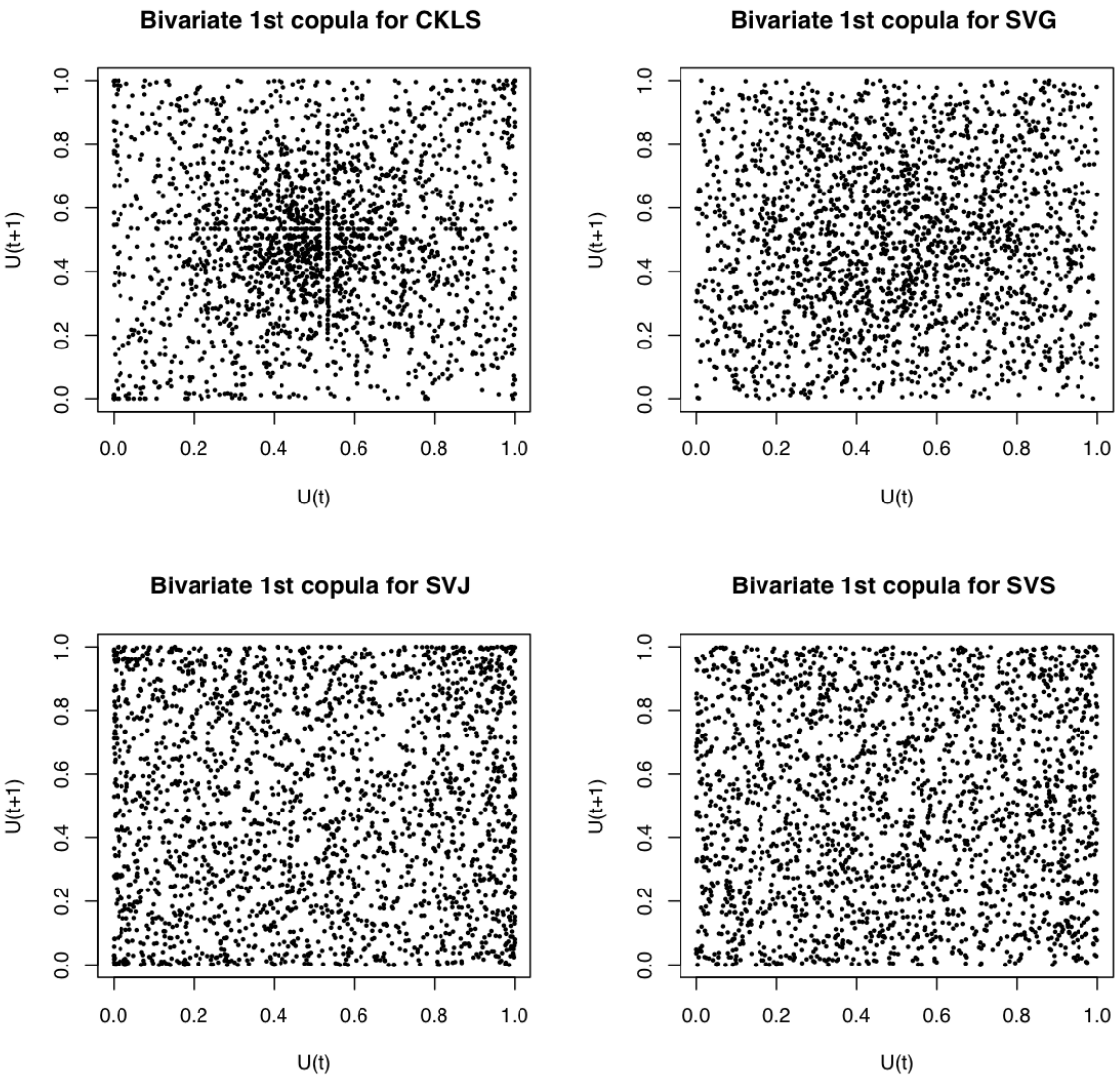

Figure 4.28: Bivariate copulas of $(U_i(t), U_i(t+1))$ for the CKLS, SVG, SVJ and SVS model respectively.

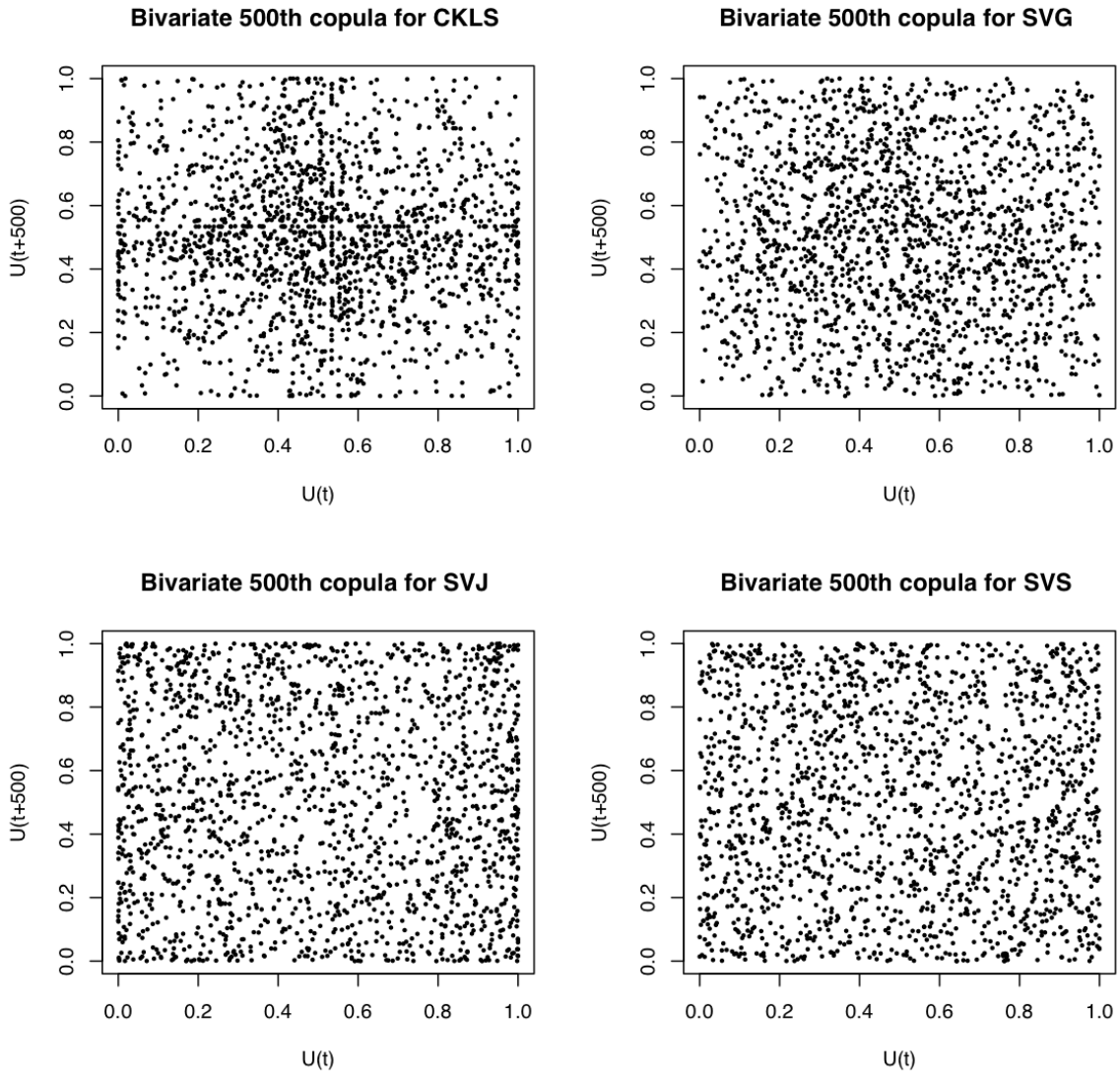

Figure 4.29: Bivariate copulas of $(U_i(t), U_i(t+500))$ for the CKLS, SVG, SVJ and SVS model respectively.

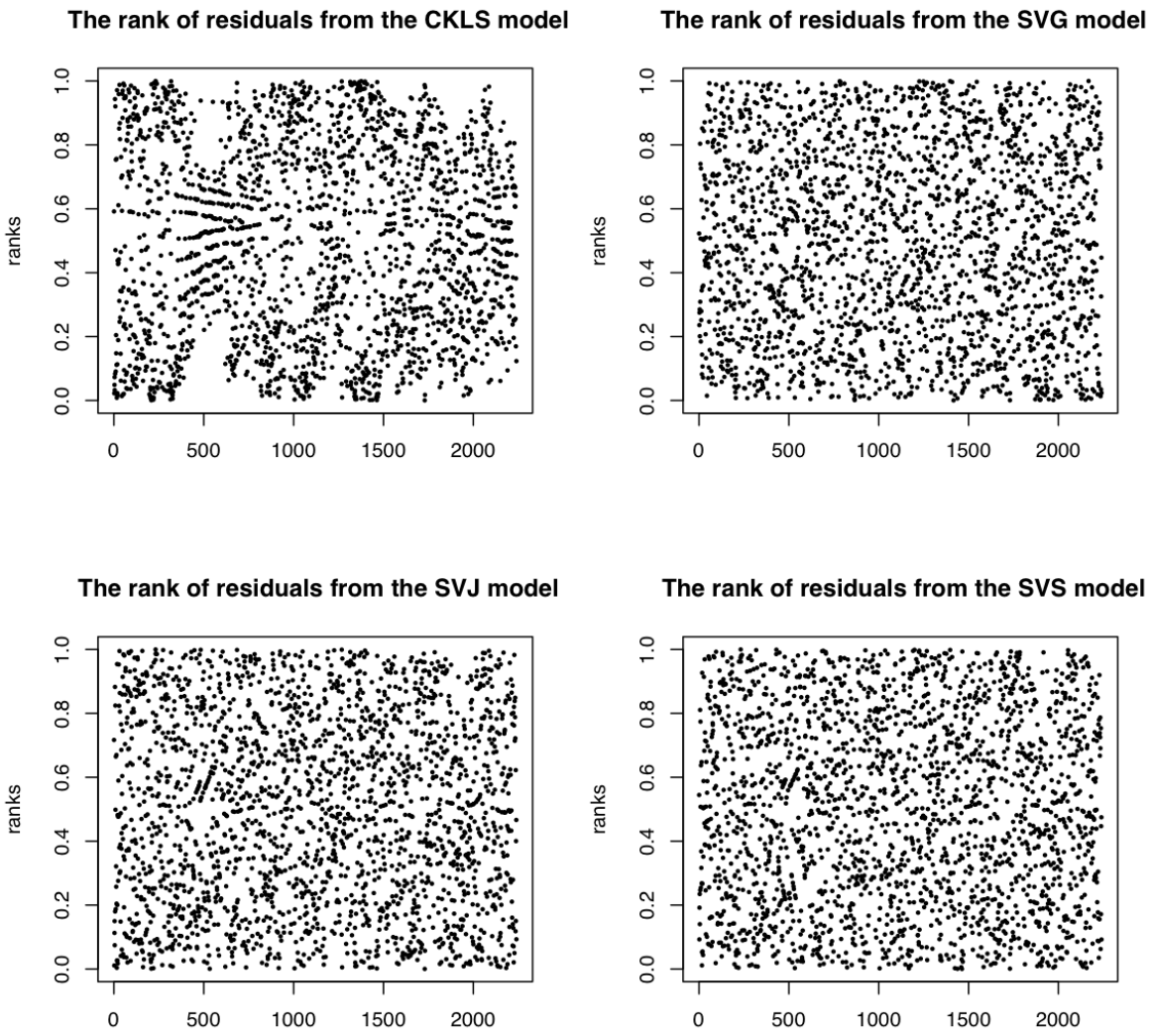

Figure 4.30: The rank of the residuals from the CKLS, SVG, SVJ and SVS model respectively.

Table 4.12: Diagnostic test of the independent copula

	CKLS	SVG	SVJ	SVS
χ^2 statistics	1306.964	341.875	327.143	106.071
p-value	0	0	0	0.295

generally decreases. For the CKLS model, the positive correlation is not as obvious as in the first copula. But the points are still highly concentrated in the middle, 28% in this case. The time lag effect to the SVG model is also not apparent. There are still less points around each corner as we can see from the top-right picture and 22% points in the center area. For the SVJ model, the number of points gathering in corners obviously decrease as we check the left-bottom plot. But there are still some points grouping in the right-top corner. The percentage of central points is dropping slightly to 11%. The points for the SVS model are uniformly distributed as previous one, although the percentage of central points is also dropping to 14%.

To prove our observation from the copula plots, we will apply the Chi-square test described as follows. First, the unit square area is divided into 10×10 parts. Let o_{ij} be the observed number of points in the i, j th segment for $i, j = 1, ..., 10$. Then the χ^2 statistic is defined as:

$$X = \sum_{i,j=1}^{10} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

where e_{ij} is the expected number in ijth segment for $i, j = 1, ..., 10$ which equals 2240/100 under the uniform hypothesis. The degrees of freedom of the Chi-square distribution is 99.

Table 4.12 reports the χ^2 statistics and p values for each model. The values of χ^2 statistics decrease dramatically from 1306.964 for the CKLS to 106.071 for the SVS model. All the models except SVS fail to pass the chi-square test. It is also worth to notice the improvement by changing the continuous volatility to the jump process. The failure of the rejected models are either due to the assumption of Normal

Table 4.13: Chi-square test for the rank

	CKLS	SVG	SVJ	SVS
χ^2 statistics	450.89	78.75	107.23	90.36
p-value	0	0.92	0.25	0.69

marginal distribution or the independence, or both.

As we know, the rank statistics of the residuals do not depend on their distributions. We can then apply a non-parametric test of the independence by checking the rank. R_i, the rank of the residual ξ_i is its position in the sequence $\xi_{(1)} < \xi_{(2)} < ... < \xi_{(n)}$. If $\{\xi_i\}$ are i.i.d., R_i will be from the discrete uniform distribution with possible values from 1 to n. Figure 4.30 presents the transformed residual ranks $U_i = (R_i - 0.5)/n$ for these models, and n is the total number of residuals. As we can see from the top left plot, the ranks for the CKLS model are obviously not uniformly distributed which concentrating around the 500th position. For the SV models, we cannot tell the differences from these plots. The corresponding Chi-square tests are presented in Table 4.13. Similar as the copula, each plot is divided by 10×10 parts, and the degree of freedom for ranks is 98. All the SV models pass the test and only the CKLS model fails.

Based on both Chi-square tests, the residuals from the CKLS model are neither Normal random variables nor independent. The residuals from both the SVG and the SVJ models are independent but not from the Normal distribution. The SVS model is the only one to pass both tests.

4.8 Bayes Factors and Model Selection

We have compared the fitting ability of each model through a series of tests about the normality and independence of corresponding residuals. The tests figure out the

misspecification of the CKLS, SVG and SVJ model. The SVS model is the most satisfied model so far. However, none of the tests we have applied so far consider whether or not the extra complexity in the SVS model is worthwhile statistically. Some kind of model selection criterion that trade off the complexity of a candidate model against its performance is necessary, such as *Akaike Information Criterion* (AIC), *Schwartz Bayes Criterion* (SBC) and Bayes factor. The Bayes factor will be a proper choice for our purpose. Since, the candidate models introduce latent processes which are considered similarly as the model parameters, neither AIC or SBC can handle such models. Another reason is that the results from the MCMC estimation can be applied to calculate the Bayes factors as we will show latter.

The Bayes factor is a model selection criterion by comparing the marginal likelihood values of the candidate models. Given the sample data $\{X\}$, the Bayes factor $B_{1,0}$ for a model M_1 against the basic model M_0 is defined as:

$$B_{1,0} = \frac{m(X|M_1)}{m(X|M_0)}$$

where $m(X|M_i)$ is the marginal likelihood function for each model M_i, $i = 0, 1$, which has the specification as below:

$$m(X|M_i) = \int \pi(X|\theta_i, M_i)\pi(\theta_i|M_i)d\theta_i$$

and θ_i is the parameter vector of model i; $\pi(X|\theta_i, M_i)$ and $\pi(\theta_i|M_i)$ are the likelihood and prior distribution respectively. Details can be found in Jeffreys (1998). Different as the test which only offers one hypothesis preferred status (the null hypothesis), there is a scale for the interpretation of Bayes factor B from Jeffreys's book:

B(1,0)	**log (B)**	**Strength of evidence**
$< 1:1$	< 0	Negative (prefers M_2)
$1:1$ to $3:1$	0 to 1.099	Barely worth mentioning
$3:1$ to $12:1$	1.099 to 2.485	Positive
$12:1$ to $150:1$	2.485 to 5.011	Strong
$> 150:1$	> 5.011	Very strong

Since the marginal probability is calculated by integrating the sampling likelihood function with respect to the prior distribution of the parameters, $\pi(\theta_i|M_i)$, the posterior information from the MCMC iterations can not be used directly in this calculation. It is worth noting that the marginal likelihood $m(X|M_i)$ can also be written as:

$$m(X|M_i) = \frac{\pi(X|\theta_i, M_i)\pi(\theta_i|M_i)}{\pi(\theta_i|X, M_i)}$$

where $\pi(\theta_i|X, M_i)$ is the joint posterior density of parameters. Therefore, the calculation is down to the estimation of the posterior distribution. The method used here to estimate $\pi(\theta_i|X, M_i)$ is based on the output of Metropolis-Hastings algorithm introduced by Chib & Jeliazkov (2001).

Their approach is based on the reversibility of the subkernel of the MH algorithm. Recall that the acceptance probability in MH is:

$$\alpha(\theta, \theta^{'}|y) = \min\left(1, \frac{L(y|\theta^{'})\pi(\theta^{'})q(\theta^{'}, \theta\|y)}{L(y|\theta)\pi(\theta)q(\theta, \theta^{'}|y)}\right),$$

where $L()$ is the likelihood for sample y, $\pi(\theta)$ is prior density of θ and $q(\theta, \theta^{'}|y)$ is the proposal density from θ to $\theta^{'}$. The subkernel is defined as $p(\theta, \theta^{'}|y) = \alpha(\theta, \theta^{'}|y)q(\theta, \theta^{'}|y)$. According to the subkernel reversibility, we have the following equation:

$$p(\theta, \theta^{'}|y)\pi(\theta|y) = p(\theta^{'}, \theta|y)\pi(\theta^{'}|y).$$

Then for any point $\theta^{'}$, we obtain:

$$\pi(\theta^{'}|y) = \frac{\int \alpha(\theta, \theta^{'}|y)q(\theta, \theta^{'}|y)\pi(\theta|y)d\theta}{\int \alpha(\theta^{'}, \theta|y)q(\theta^{'}, \theta|y)d\theta},$$

which can also be written as:

$$\pi(\theta^{'}|y) = \frac{E_1(\alpha(\theta, \theta^{'}|y)q(\theta, \theta^{'}|y))}{E_2(\alpha(\theta^{'}, \theta|y))},$$

where E_1 is the expectation respect to $\pi(\theta|y)$ and E_2 is the one respect to $q(\theta^{'}, \theta|y)$.

For the model with n parameters and latent process $\{z\}$, the marginal likelihood is down to calculate the joint posterior distribution and which can be written as:

$$\pi(\theta_1^{'}, \theta_2^{'}, ..., \theta_n^{'}|y) = \pi(\theta_1^{'}|y)\pi(\theta_2^{'}|y, \theta_1^{'})...\pi(\theta_n^{'}|y, \theta_1^{'}, \theta_2^{'}, ..., \theta_{n-1}^{'}).$$

Table 4.14: Bayes Factors: $logB_{1,0}$

	CKLS	SVG	SVJ	SVS
CKLS	0	-581.1	-633.5	-843.4
SVG	581.1	0	-52.4	-262.4
SVJ	633.5	52.4	0	-209.9
SVS	843.4	262.3	209.9	0

$\pi(\theta_i^{'}|y,\theta_1^{'},\theta_2^{'},...,\theta_{i-1}^{'})$ can also be calculated from the local reversibility property. Let $\Theta_{i-1^-} = (\theta_1,...,\theta_{i-1})$ and $\Theta_{i+1^+} = (\theta_{i+1},...,\theta_n)$. We will obtain the following approximation from MCMC:

$$\hat{\pi}(\theta_i^{'}|y,\Theta_{i-1^-}^{'}) = \frac{M^{-1}\sum_{g=1}^{M}\alpha\left(\theta_i^{(g)},\theta_i^{'}|y,\Theta_{i-1^-}^{'},\Theta_{i+1^+}^{(g)},z^{(g)}\right)q\left(\theta_i^{(g)},\theta_i^{'}|y,\Theta_{i-1^-}^{'},\Theta_{i+1^+}^{(g)},z^{(g)}\right)}{J^{-1}\sum_{j=1}^{J}\alpha\left(\theta_i^{'},\theta_i^{(j)}|y,\Theta_{i-1^-}^{'},\Theta_{i+1^+}^{(j)},z^{(j)}\right)}.$$

In the nominator, the simulation of $\{\theta_i^{(g)},...,\theta_n^{(g)}\}$ and $\{z^{(g)}\}$ are generated from $\pi(\Theta_{i^+},z|y,\Theta_{i-1^-}^{'})$, $g=1,...,M$. The simulation of $\{\theta_{i+1}^{(j)},...,\theta_n^{(j)}\}$ and $\{z^{(j)}\}$ in the dominator are generated from $\pi(\Theta_{i+1^+},z|y,\Theta_{i^-}^{'})$, $j=1,...,J$. $\{\theta_i^{(j)}\}$ are simulated from the proposal distribution $q(\theta_i^{'},\theta_i|y,\Theta_{i-1^-}^{'},\Theta_{i+1^+}^{(j)},z^{(j)})$. Details can be found in Chib & Jeliazkov (2001).

The value of marginal log-likelihood function for each model is listed at the bottom of corresponding parameter estimation table: from Table 4.7 to Table 4.11. The results of log-Bayes factors are presented in Table 4.14 where we put the basis model on the top row and list the name of candidate in left border column. There is very strong evidence that the SV model outperforms the CKLS model since the Bayes factor is much larger than the criterion value 5. There is also a very strong evidence that the single jump process performs better than the Gaussian process. Although we couldn't observe such advancement directly though our previous tests, the Bayes factor for the SVJ model against the SVG model shows such evidence. In conclusion , we have strong evidence for the SVS model to back up our earlier conclusions.

Table 4.15: Parameter estimates of GARCH(p,q) models for the weekly 3-month U.S. T-bill

Parameters	GARCH(1,1)		GARCH(2,1)		GARCH(1,2)	
	Mean	*Std.*	*Mean*	*Std.*	*Mean*	*Std.*
α $\times 10^{-5}$	4.96	0.08	3.67	0.08	4.95	0.08
β $\times 10^{-4}$	-3.19	0.18	-1.10	0.20	-3.07	0.18
γ	0.902	0.001	0.992	0.001	0.887	0.001
μ $\times 10^{-5}$	7.123	0.074	11.11	0.011	6.522	0.063
a_1	0.606	0.001	0.409	0.002	0.563	0.001
b_1	0.375	0.001	0.364	0.001	0.380	0.001
a_2	–	–	0.205	0.002	–	–
b_2	–	–	–	–	0.061	0.001
Marginal Log-Likelihood		11335.6		11342.9		11348.8

4.9 Volatility Analysis

This section provide the estimated volatility processes from the SVG, SVJ and SVS model. For comparison, we also present the GARCH model alongside SV models to help us examine the performance of different driven processes.

4.9.1 Volatility process from GARCH model

It is well known that an class of models used to capture the volatility changing in both asset models and interest rate models is autoregressive conditional heteroscedastic models (ARCH) or generalized ARCH models (GARCH). These kinds of models were firstly introduced by Engle (1982)(ARCH), and latter modified in Bollerslev (1986) and in Taylor (1986) independently (the GARCH model). They become very popular in both academic and industry thereafter. Unlike the SV models, there are no new factors introduced to explain the volatility process. The volatility described by GARCH models is merely a function of residuals and previous variances. Then calibrations of GARCH model are more straightforward than SV model. Because of this we will compare the results of our models against the GARCH model as a benchmark.

The GARCH model used in this section has the general specification as below:

$$\begin{aligned} \Delta r_t &= (\alpha + \beta r_t)\Delta t + r_t^{\gamma}\varepsilon_t \\ \varepsilon_t &\sim N(0, h_t^2) \\ h_t^2 &= w + \sum_{i=1}^{p} a_i h_{t-i}^2 + \sum_{j=1}^{q} b_i \varepsilon_{t-j}^2 \end{aligned}$$

where α, β, γ, w, $\{a_i\}_{i=1}^{p}$ and $\{b_i\}_{i=1}^{q}$ are constant parameters, $N(0, h_t^2)$ is normal distribution with mean 0 and variance h_t^2. It is a discrete form of CKLS model incorporated with GARCH(p, q) model. In the GARCH part, $\{a_i\}_{i=1}^{p}$ and $\{b_i\}_{i=1}^{q}$ are the weighting factors, $w > 0$, $a_i > 0$ and $b_j > 0$ for $i = 1, ..., p$ and $j = 1, ..., q$. The number of time lags p, q for the past observations of residuals and variance respectively can be determined by a series of information criteria, such as the AIC, BIC, HQC (Hannan-Quinn Information Criterion) and Bayes factor.

Table 4.15 presents the estimation results for several GARCH models. The historical data used to fit them are the same set as we used for CKLS and SV models. The estimation method is also MCMC. The sample length for each parameter is $600,000$ with the first $100,000$ iterations as burn-in. At the bottom of the table, the value of marginal log-likelihood for each model is listed respectively. Obviously, according to the Bayes factor, GARCH(1,2) performs better than the other two which is the same conclusion as Andersen & Lund (1997) whose selection is based on HQC. However, it is still less the marginal log-likelihoods for the SVG, SVJ and SVS models. The estimated volatility process, $\sigma_1(t) = r_t^{\gamma} h_t$, from GARCH(1,2) is plotted against the paths from the stochastic volatility models which would enable us to investigate the improvement of introduce extra factor in volatility. Besides we also present the comparison of underlying pure volatility effects, $v(t)$ which equals to h_t for the GARCH model.

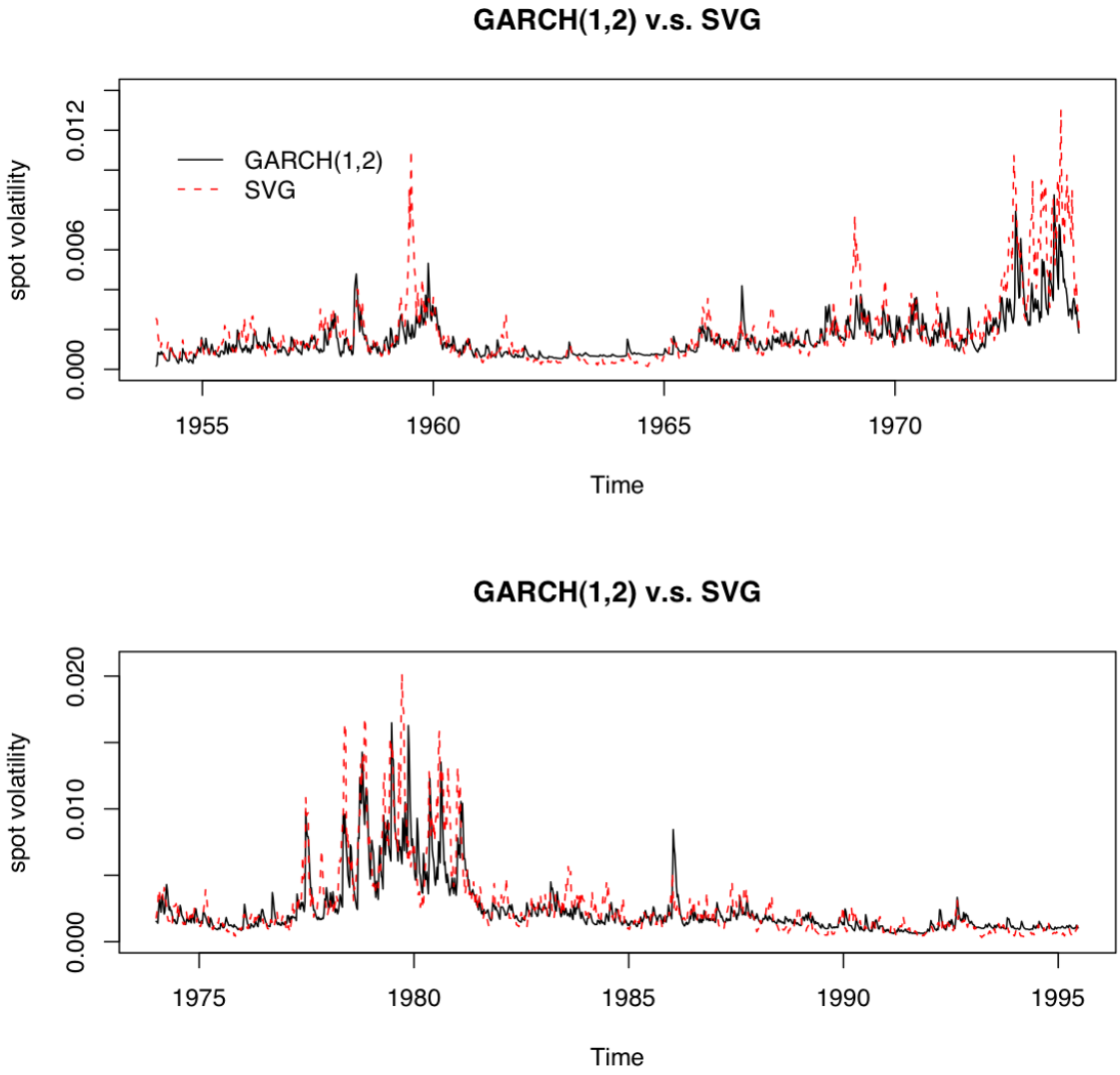

Figure 4.31: The estimated volatility $\sigma_2(t) = \sigma_r r_t^{\gamma} e^{h_t/2}$ paths from the GARCH model and the SVG model, where $\gamma = 0.714$ in the SVG model and $\gamma = 0.887$ in the GARCH(1,2) model.

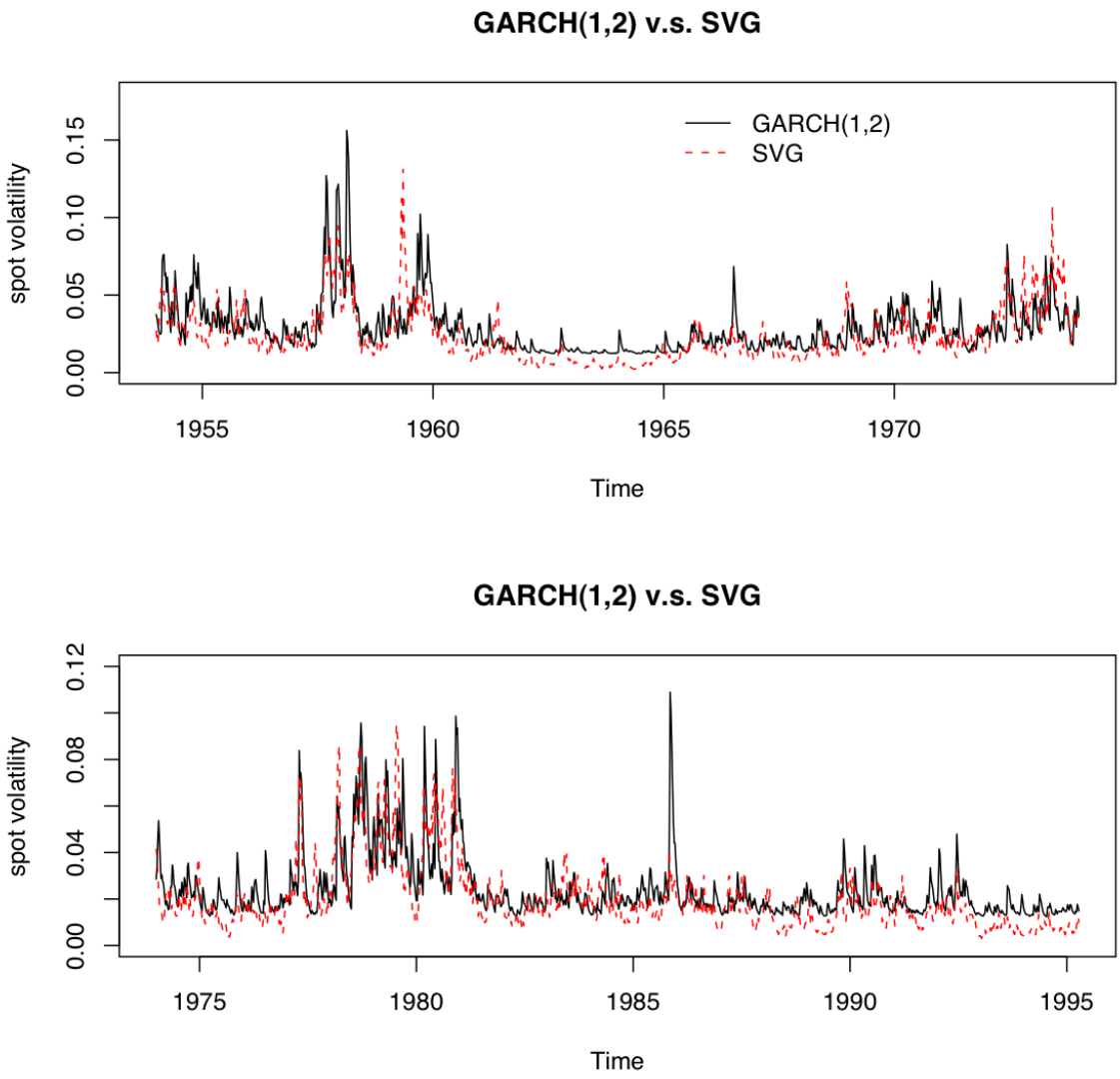

Figure 4.32: The underlying volatility $v_2(t) = \sigma_r e^{h_t/2}$ paths from the GARCH model and the SVG model.

4.9.2 Volatility process from the SVG model

In Figure 4.31, we compare the estimated volatility processes , $\sigma_2(t) = \sigma_r r_t^{\gamma} e^{h_t/2}$, from the SVG and the GARCH model in four periods, 1954-1964, 1964-1974, 1974-1984 and 1984-1997. The underlying volatility $v_2(t) = \sigma_r e^{h_t/2}$ for the SVG model against $v_1(t)$ the one from the GARCH(1,2) model is presented in Figure 4.32 with the same time decomposition.

There are two points worthy of mention. Firstly, the process $\sigma_2(t)$ from the SVG model is above the the GARCH model over some periods, especially for the highly volatile time such as 1978-1981 which is one of the period that the interest rate reached a high level. One of the reasons is that the leverage effect. As we know the γ in the SVG model drops from 0.887 (in GARCH(1,2), Table 4.15) to 0.714 (Table 4.9) which will make the estimated volatility bigger when the underlying volatility processes $v(t)$ from two models are near to each other. Secondly, as we can see from Figure 4.32, although $v_2(t)$ representing by the dash line is below $v_1(t)$, the solid line for most of the periods, there are still some periods when $v_2(t)$ is higher than $v_1(t)$ again for the highly volatile time. The SVG model tries to capture the large variations in interest rates not only by higher leverage effect but also through sudden bigger movements in the underlying volatility process. Obviously, such movements are inconsistent with the normal assumption of the volatility driven process and it is another evidence of the misspecification of the SVG model. This drawback of the SVG model is also pointed out by Andersen *et al.* (2001) and Eraker *et al.* (2000).

4.9.3 Volatility process from the SVJ model

For the SVJ model, the jump process is presented in Figure 4.33, and the estimated volatility path $\sigma_3(t) = r_t^{\gamma} h_t^{1/2}$ is in Figure 4.34 comparing with $\sigma_1(t)$ from the GARCH(1,2) model. Also the underlying process $v_3(t)$ is plotted in Figure 4.35 comparing with $v_1(t)$ from GARCH(1,2).

There are several points need to mention for the comparison between the volatility

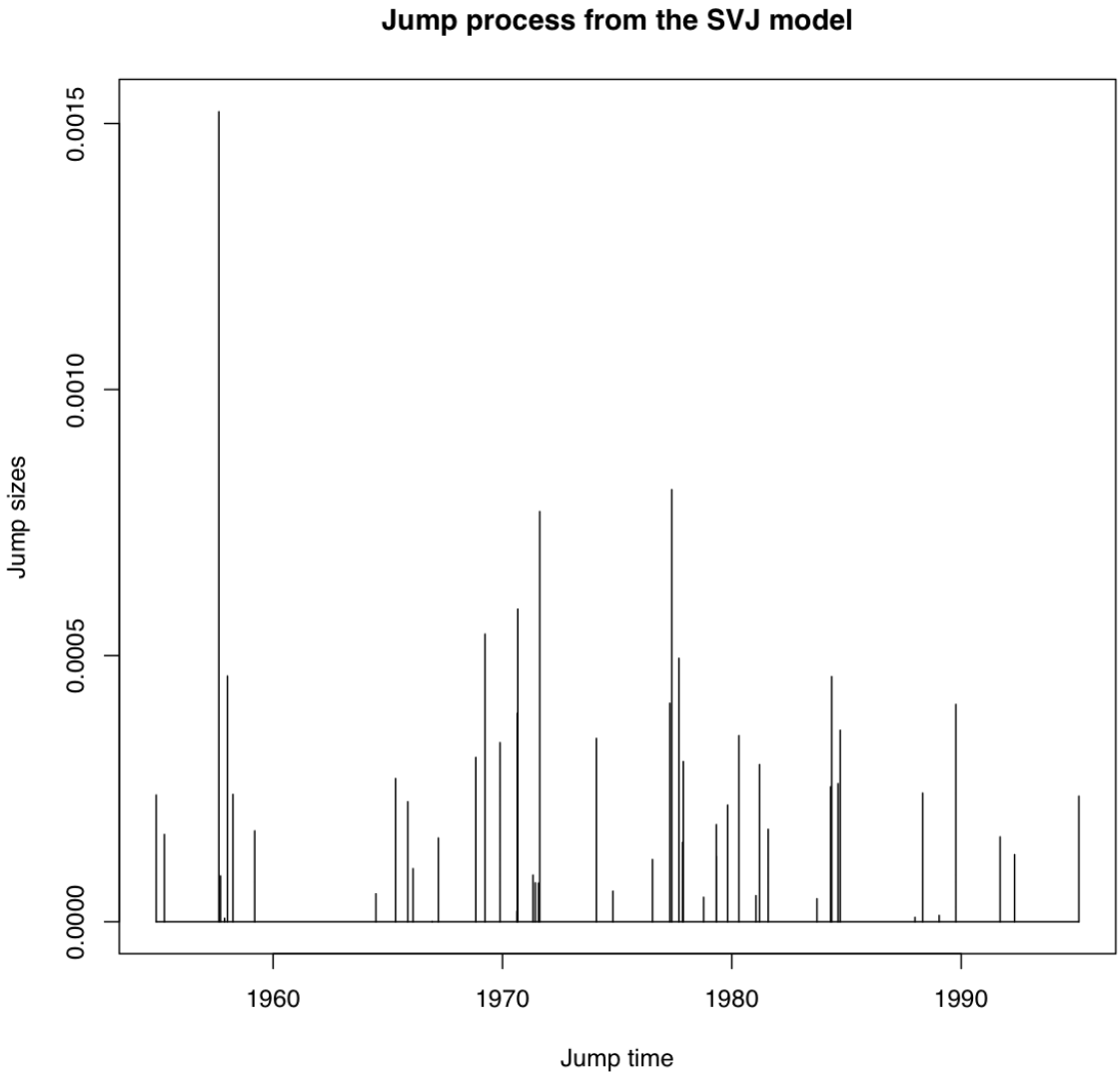

Figure 4.33: The jump points from the SVJ model.

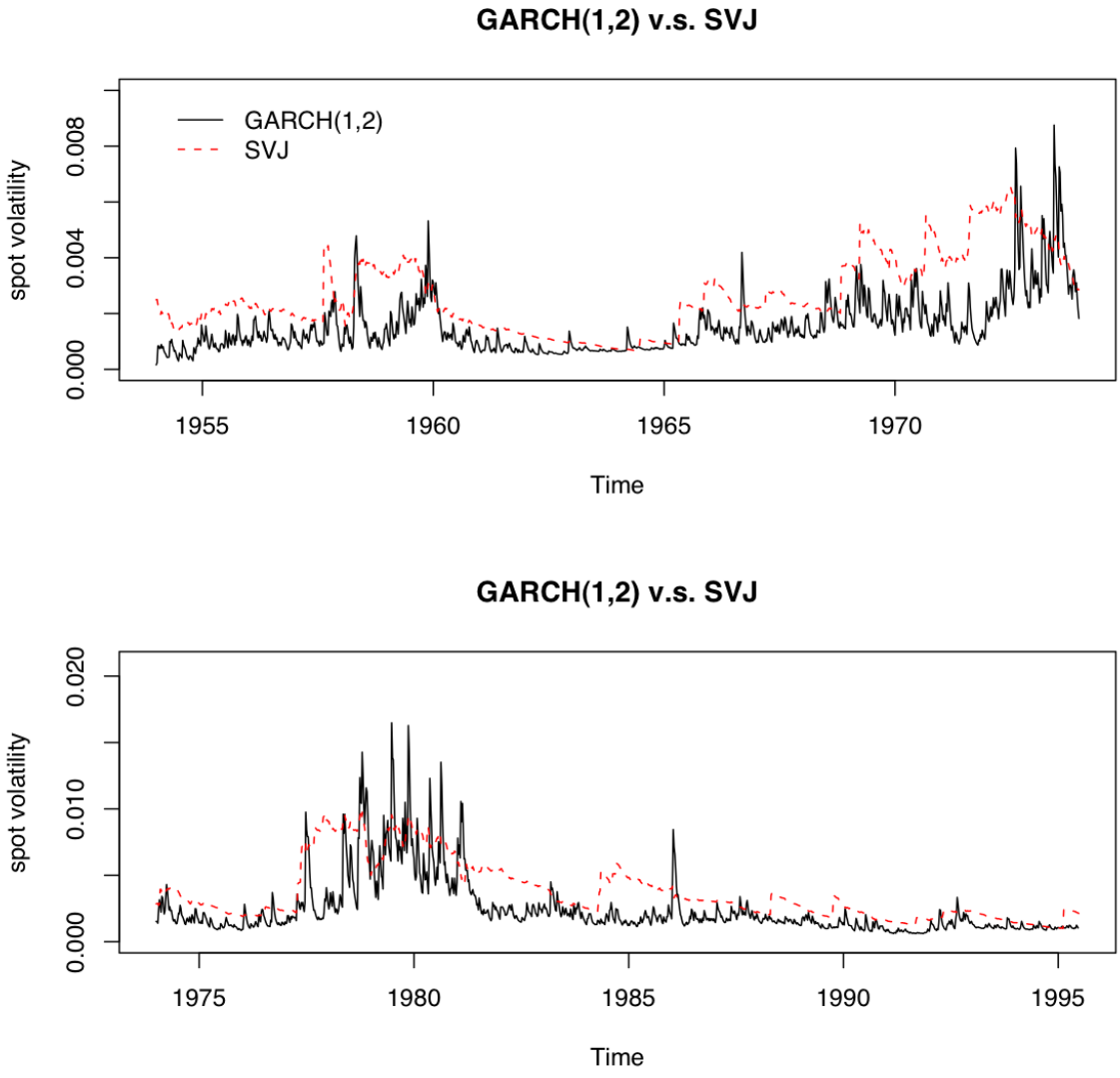

Figure 4.34: The estimated volatility $\sigma_3(t) = r_t^{\gamma} h_t^{1/2}$ paths from the GARCH model and the SVJ model, where $\gamma = 0.642$ in the SVJ model and $\gamma = 0.887$ in the GARCH(1,2) model.

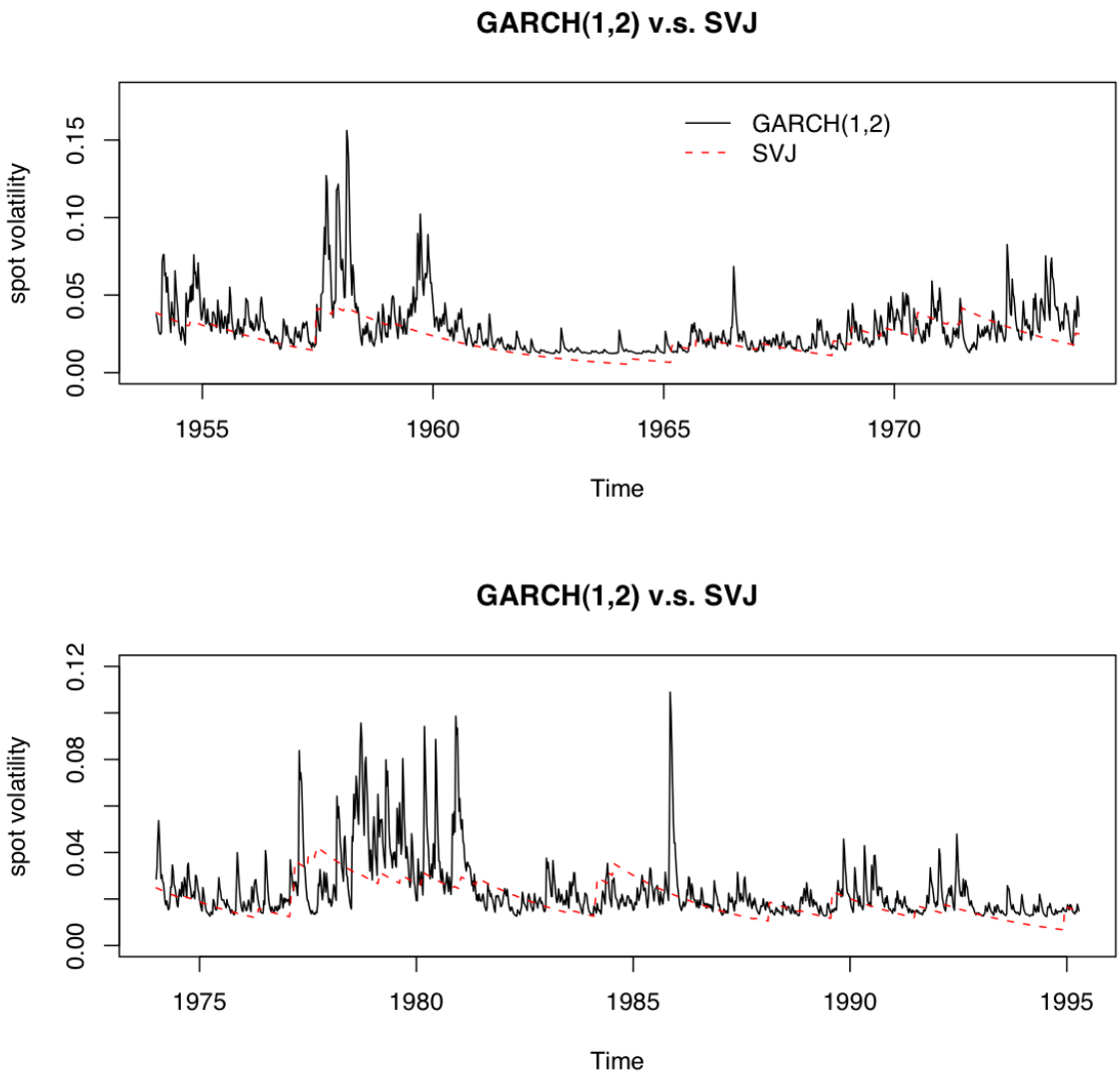

Figure 4.35: The underlying volatility $v_3(t) = h_t^{1/2}$ paths from the GARCH model and the SVJ model.

processes from the SVJ model and the GARCH model. Firstly, there are only about 1.23 jumps per year on average according to the SVJ model. As we can observe from Figure 4.33, the jumps don't happen uniformly over these years but concentrate over some time, 1978-1981 for example, whereas there are no jumps over 1960-1963 at all. with single jump process. These indicate that the single jump process in the volatility part is insufficient to capture all the variations in volatility.

Secondly, shown in Figure 4.35, the underlying volatility process $v_3(t)$ is very close to the bottom of $v_1(t)$. Due to the model structure (increase by jump and decaying exponentially), $v_3(t)$ is much smoother than $v_1(t)$. γ in the SVJ model is also decreasing to 0.642. With such high level effect, the estimated volatility process from the SVJ model, $\sigma_3(t)$ is covering $\sigma_1(t)$ for most of the period, see Figure 4.34.

Thirdly, when we compare $v_3(t)$ the underlying volatility process with the one from the SVS model, shown in Figure 4.38, we can find another evidence of the misspecification of the SVJ model. The SVJ model try to cover more variation with few higher jumps and decaying more slowly than the SVS model, see the period 1958-1964 for example. Therefore, the SVJ model would overestimate the variation of volatility for most of the time.

4.9.4 Volatility process from the SVS model

The jump processes for each part of the SVS model are presented in Figure 4.36. We also plotted the underlying volatility processes for each part, $v_{4,1}(t) = h_1(t)^{0.5}$ the one with high decay rate, and $v_{4,2}(t) = h_2(t)^{0.5}$ with low rate. Figure 4.39 and Figure 4.40 provide the estimated volatility process $\sigma_4(t) = h(t)^{0.5} r_t^{\gamma}$ and the underlying process $h(t)^{0.5}$ against the corresponding processes from GARCH(1,2) model in decomposed time periods.

Jump in the process with low decay rate still happens rarely , about 1.5 per year,

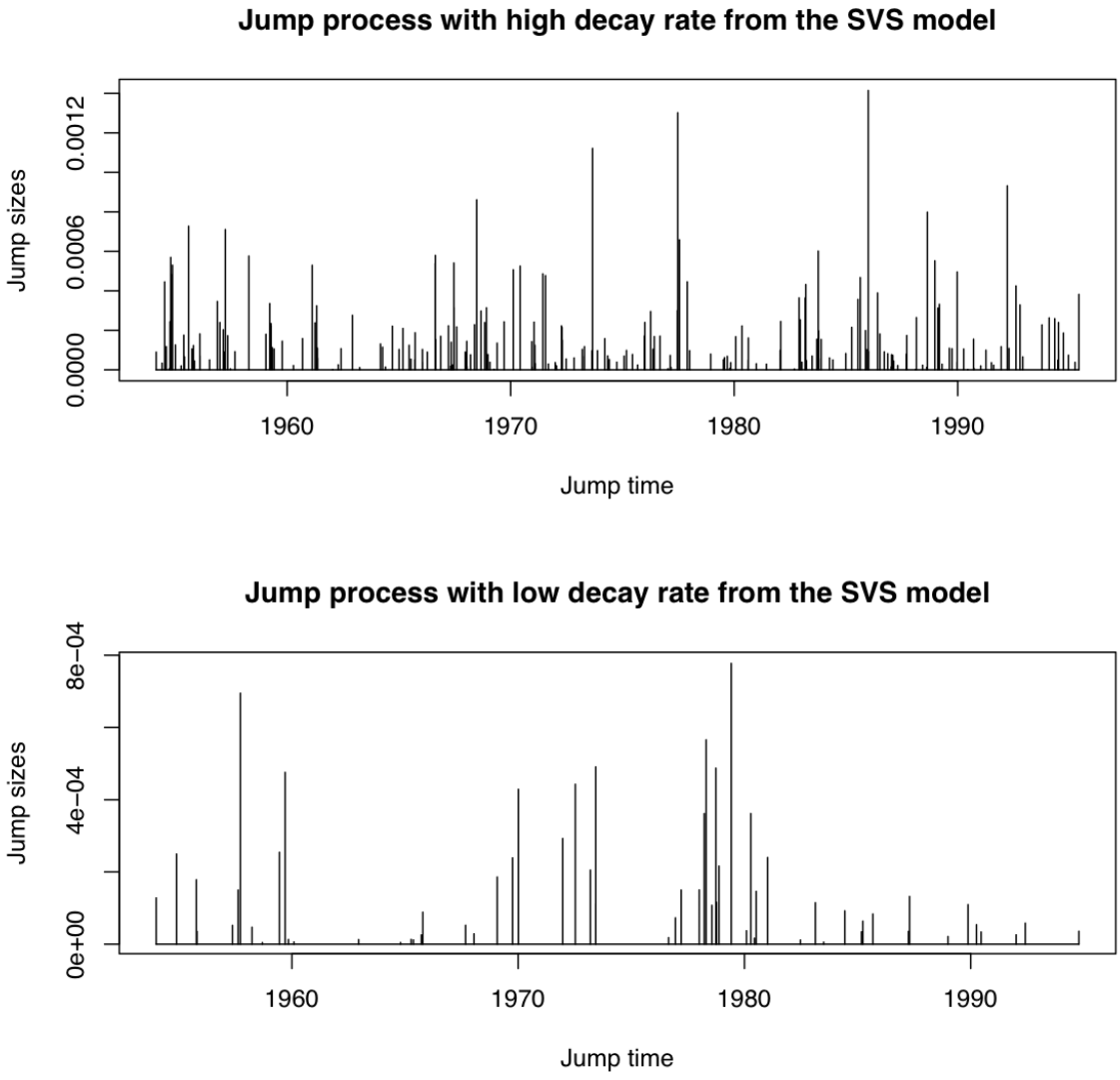

Figure 4.36: The jump points from the SVS model.

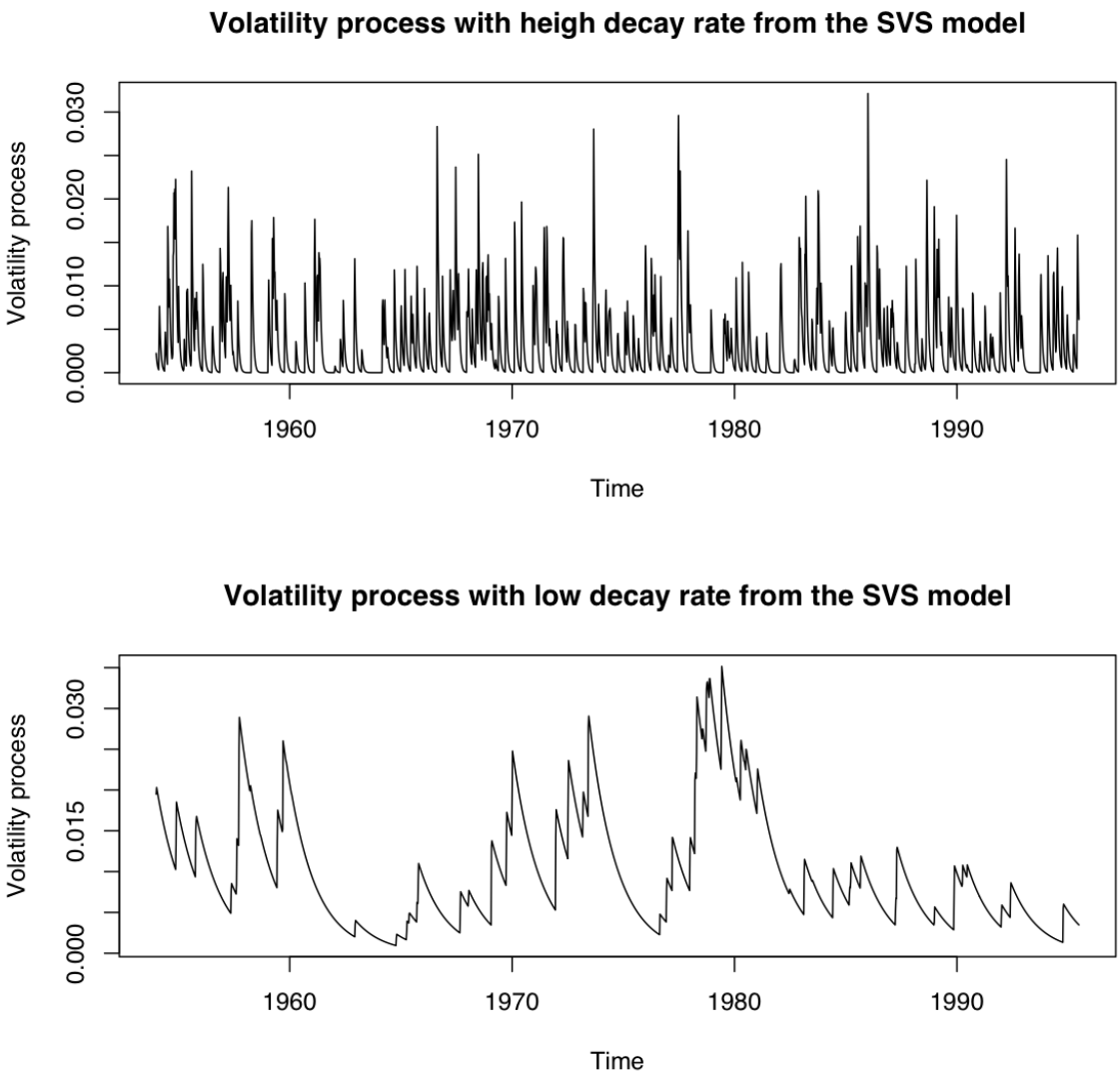

Figure 4.37: The driven random factor from the SVS model, top: $v_{4,1}(t) = h_1(t)^{0.5}$; bottom: $v_{4,2}(t) = h_2(t)^{0.5}$.

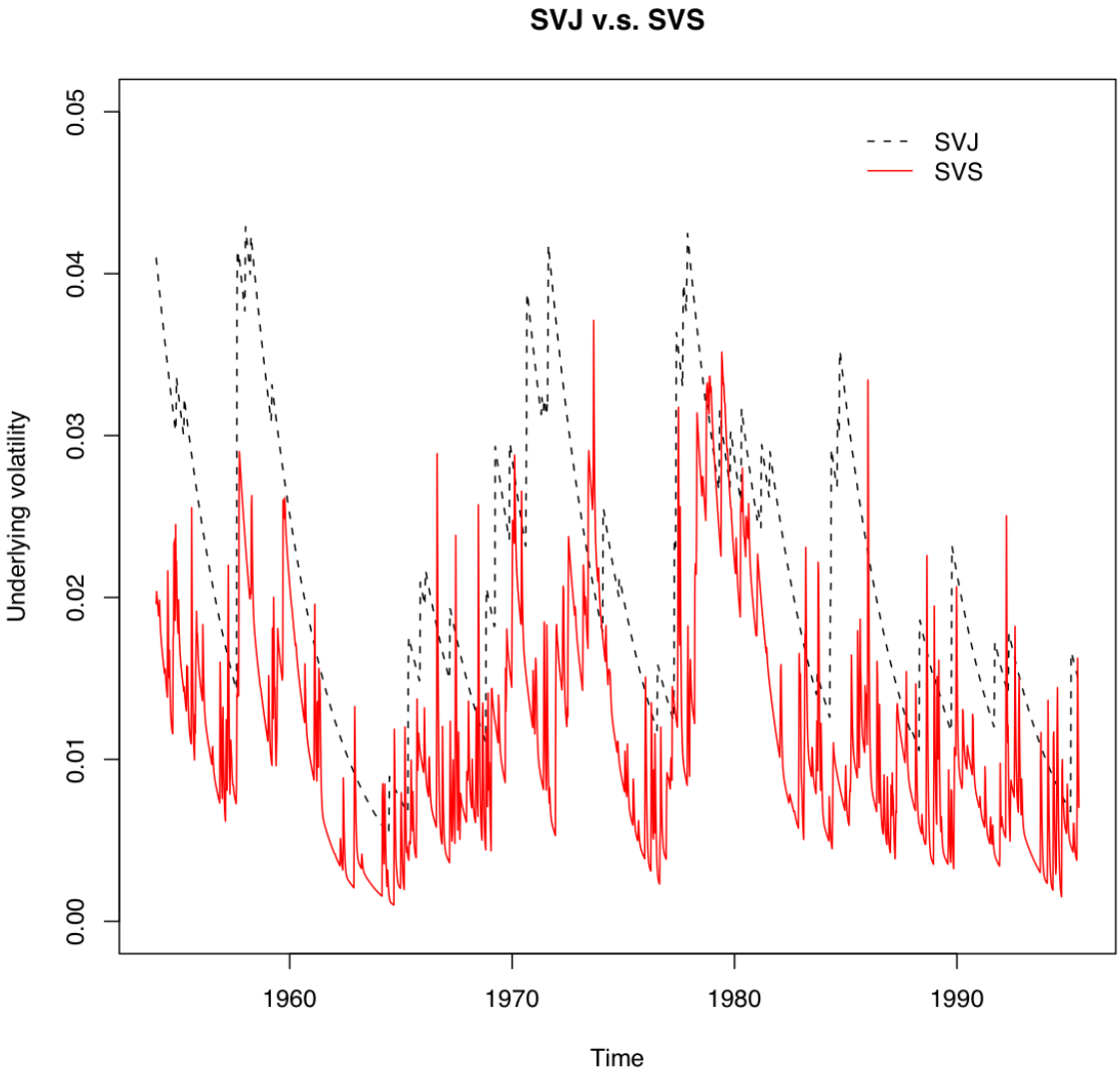

Figure 4.38: The comparison of driven random factors from the SVJ and SVS model.

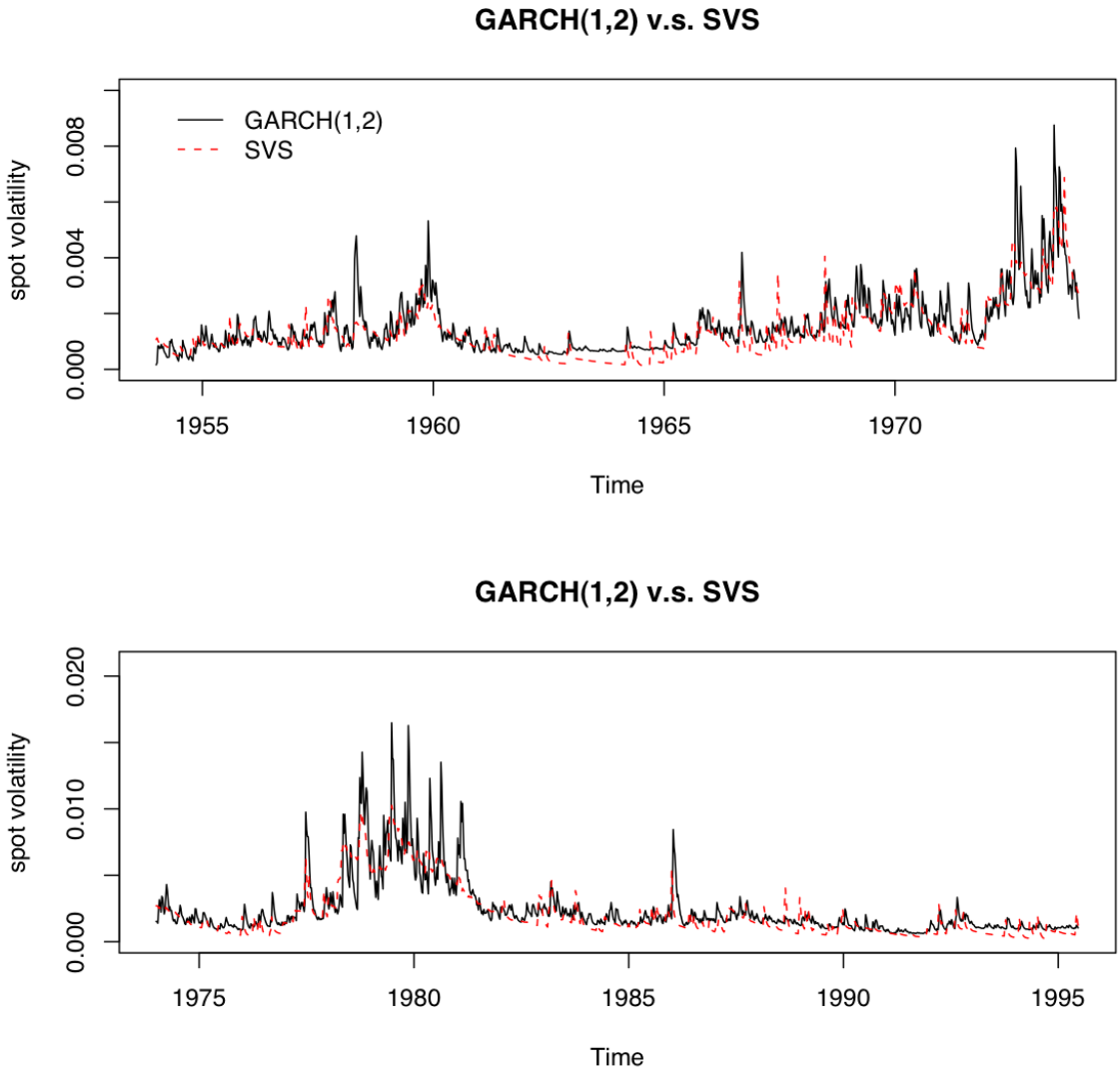

Figure 4.39: The estimated volatility $\sigma_4(t) = h(t)^{0.5} r_t^{\gamma}$ paths from the GARCH model and the SVS model, where $\gamma = 0.672$ in the SVG model and $\gamma = 0.887$ in the GARCH(1,2) model.

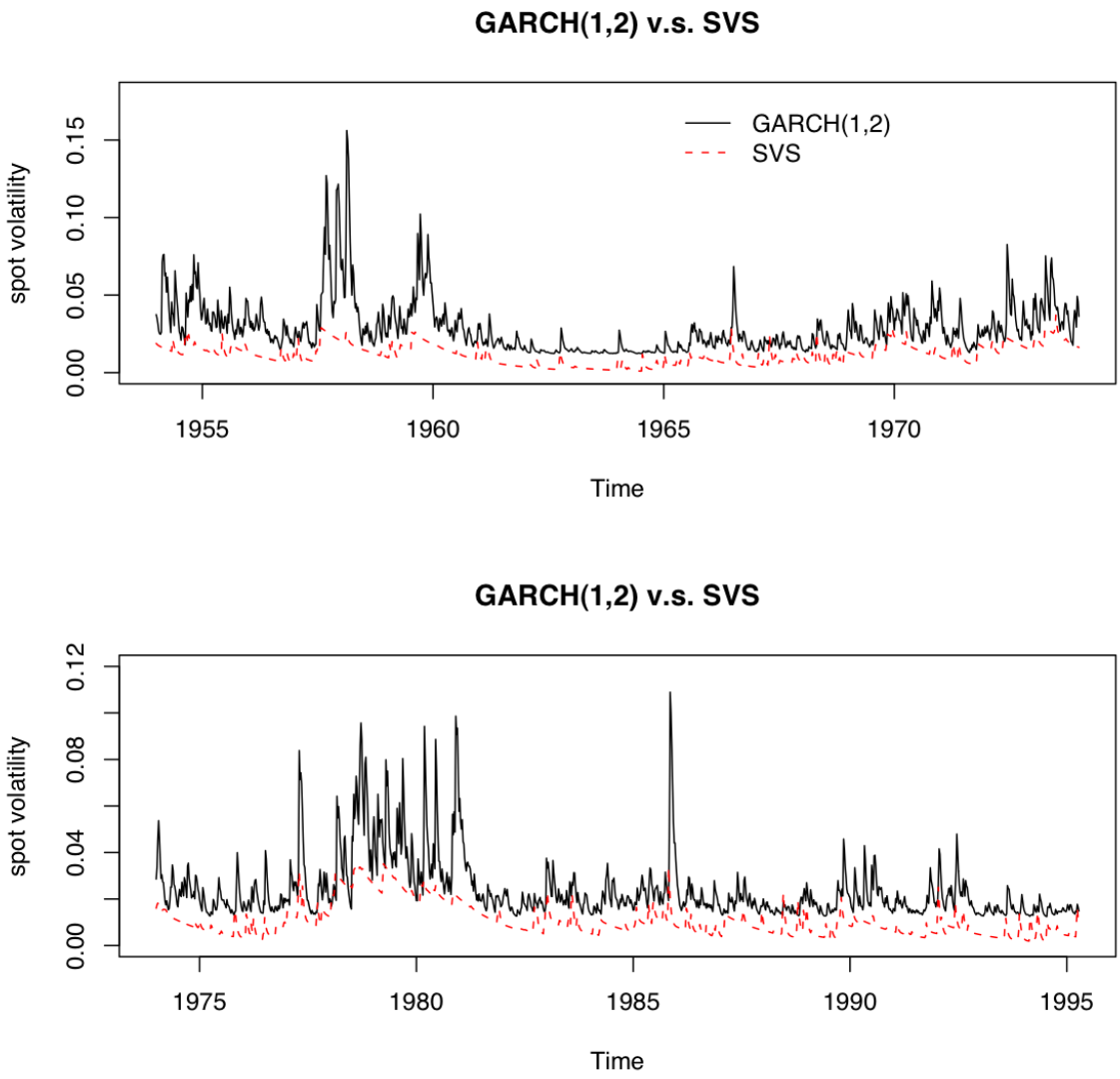

Figure 4.40: The underlying volatility $v_4(t) = h(t)^{0.5}$ paths from the GARCH model and the SVS model.

and its pattern is similar as the one from the SVJ model, see the bottom plot of Figure 4.36. The jump sizes are generally smaller than the SVJ ones. With the jump process with higher decay rate, more jumps can be captured by the SVS model. In $h_2(t)$, there are about 5.3 more jumps per year. As a result, the superposition structure make the SVS model more flexible and enable it to capture the variation happens during the decaying period, see Figure 4.38.

In Figure 4.40, the underlying volatility process $v_4(t)$ from the SVS model is going under $v_1(t)$ for almost all the observation period but much lower the one from the SVJ model. This is due to the smaller jump sizes and faster decaying rate comparing with the SVJ model. The estimated volatility process from the SVS model $\sigma_4(t)$, shown in the Figure 4.39, is quite match to $\sigma_1(t)$, since the leverage effect is more significant than the GARCH model with γ decreasing to 0.672.

4.10 Conclusion

In this chapter, we have presented the estimation results of the MCMC algorithm developed in the Chapter 3 for each model. We check the describing abilities of these models by examine the corresponding residuals. If the candidate can fully explain the historical observation, the residuals should be i.i.d. random variables with Normal marginal distribution. As we have shown, the SVS model is the only candidate which is not rejected by the statistical tests we considered. For the SV models, we also test their ability in describing the latent volatility process by comparing with GARCH model. We choose the GARCH(1,2) model as the benchmark. The failure of the SVG model is the inability of capturing the bigger movements. The SVJ model fails to capture most of various period of the latent process. While the composition structure offers the SVS model more flexibility in capturing the volatility process. Therefore, the SVS model is the most suitable model in explaining the historical short rate process.

Chapter 5

Pricing Interest Rate Products

5.1 Introduction

The models we have studied in the previous chapters describe the random evolutions of the short rate and the stochastic volatility under the canonical measure (the real world measure). The parameters estimated from the historical interest rates are therefore the real world parameters. In order to apply such models in pricing, we should change the canonical measure to some kinds of *pricing measure* and figure out such SDEs under the new measure.

The *change of measure* is a methodology to construct a new dynamic process on the same set of random path by assigning new probabilities to same events. As we mentioned in section 1.5.2, we can find the arbitrage-free price for a financial instrument if there is an equivalent martingale measure. Under such a measure, the relative price of the instrument, which is the price divided by the corresponding numeraire, will be a martingale. If such an equivalent martingale measure is unique given the numeraire, we say the market is *complete*, otherwise it is a *incomplete market*, (see, for example, Brigo & Mercurio (2006)). The equivalent martingale is called *risk-neutral measure* if we take the money-market account B_t as the numeraire.

In the following sections, we will first prove the existence of equivalent martingale measure for these models. Then the derivative prices can be calculated through

Monte Carlo simulations even without knowing analytical pricing formulas.

5.2 Measure Changing

5.2.1 CKLS Model Under the Risk Neutral Measure

In this and following subsections, we will consider the risk neutral measure and assign the money-market account B_t which is defined as

$$dB_t = r_t B_t dt, \tag{5.84}$$

as the numeraire.

In the market where the short rate process is described by the CKLS model, the value of interest rate instrument, V, is therefore a function of both time and short rate. Accompanied with Equation 1.20 and the Itô lemma, the SDE of V_t can be written as:

$$dV_t = \mu_V V_t dt + \sigma_V V_t dW_t \tag{5.85}$$

where

$$\begin{aligned} \mu_V &= V_t^{-1}\left(\frac{\partial V}{\partial t} + \frac{\partial V}{\partial r_t}(\alpha + \beta r_t) + \frac{1}{2}\frac{\partial^2 V}{\partial r_t^2} r_t^{2\gamma}\sigma_r^2\right), \\ \sigma_V &= V_t^{-1}\left(\frac{\partial V}{\partial r_t} r_t^{\gamma}\sigma_r\right), \end{aligned}$$

and W_t is a Brownian motion under $\mathbb{P}$-measure. The SDE of the discounted value of such instrument $\tilde{V}_t$ is given by:

$$d\tilde{V}_t = \tilde{V}_t(\mu_V - r_t)dt + \tilde{V}_t\sigma_V dW_t. \tag{5.86}$$

According to the *Girsanov Theorem* (Theorem 1.1), the equivalent measure where the discounted value of any instruments are martingale exists if we can find a stochastic process λ_t which satisfying the *Novikov condition*. For the single factor model,

the *Radon-Nikodym derivative* $\frac{d\mathbb{Q}}{d\mathbb{P}} = \rho_t$ is given by:

$$d\rho_t = \rho_t \lambda_t dW_t,$$

and $\rho_0 = 1$, or more generally as

$$\rho_t = \exp\{\int_0^t \lambda_s dW_s - \frac{1}{2}\int_0^t \lambda_s^2 ds\}.$$

We can figure out the λ_t through the SDE of $\rho_t \tilde{V}_t$ which is:

$$d\rho_t \tilde{V}_t = \rho_t \tilde{V}_t(\mu_V - r_t + \sigma_V \lambda_t)dt + \rho_t \tilde{V}_t(\sigma_V + \lambda_t)dW_t.$$

Since $\rho_t \tilde{V}_t$ is a $\mathbb{P}$-martingale, the drift part of the above SDE should be zero. Therefore, λ_t is given by:

$$\lambda_t = -\frac{\mu_V - r_t}{\sigma_V},$$

which is also known as the *market price of risk*.

Under $\mathbb{Q}$-measure, the SDE for the discounted value $\tilde{V}$ is:

$$d\tilde{V}_t = \tilde{V}_t \sigma_V dW_t^{\mathbb{Q}},$$

where $W_t^{\mathbb{Q}} = W_t - \int_0^t \lambda_s ds$ is a $\mathbb{Q}$-measure Brownian motion . The CKLS model has the following specification under $\mathbb{Q}$:

$$dr_t = (\alpha + \beta r_t - \sigma r_t^{\gamma} \lambda_t)dt + \sigma r_t^{\gamma} dW_t^{\mathbb{Q}}. \tag{5.87}$$

We will apply the Monte Carlo simulation in later section based on the above SDE.

5.2.2 SVG Model Under the Risk Neutral Measure

For the stochastic volatility models, the value of interest rate instruments, V_t are functions of not only time and short rate but also the random factor in volatility part. If the random volatility is driven by a Gaussian process such as the SVG model, given the Equation (2.24), the SDE of V_t can be written as:

$$dV_t = \mu_{V_t} V_t dt + \sigma_{V_t,1} V_t dW_{1,t} + \sigma_{V_t,2} V_t dW_{2,t} \tag{5.88}$$

where $W_{1,t}$ and $W_{2,t}$ are two independent Brownian motions under the measure $\mathbb{P}$, and

$$\begin{aligned}
\mu_{V_t} &\equiv V_t^{-1}\left(\frac{\partial V_t}{\partial t}+\frac{\partial V_t}{\partial r_t}(\alpha+\beta r_t)-\frac{\partial V_t}{\partial h_t}\mu h_t+\frac{1}{2}\frac{\partial^2 V_t}{\partial r_t^2}r_t^{2\gamma}\sigma_r^2 e^{h_t}+\frac{1}{2}\frac{\partial^2 V_t}{\partial h_t^2}\sigma_h^2\right),\\
\sigma_{V_t,1} &\equiv V_t^{-1}\left(\frac{\partial V_t}{\partial r_t}r_t^{\gamma}\sigma_r e^{1/2h_t}\right),\\
\sigma_{V_t,2} &\equiv V_t^{-1}\left(\frac{\partial V_t}{\partial h_t}\sigma_h\right).
\end{aligned}$$

We will apply the same method as we used for the CKLS model to find the equivalent martingale measure $\mathbb{Q}$. For the SVG model, the Radon-Nikodým derivative ρ_t would have the following form:

$$d\rho_t = -\rho_t\lambda_{1,t}dW_{1,t}-\rho_t\lambda_{2,t}dW_{2,t} \tag{5.89}$$

where $\lambda_{i,t}$, $i=1,2$ is a previsable process satisfying the Novikov condition.

Accompanied with both Equation (5.88) and (5.89), we are able to deduce the SDE for the product of discounted value, $\tilde{V}_t$, and ρ_t:

$$\begin{aligned}
d\tilde{V}_t\rho_t &= \rho_t\tilde{V}_t(\mu_{V_t}-r_t+\lambda_{1,t}\sigma_{V_t,1}+\lambda_{2,t}\sigma_{V_t,2})dt\\
&\quad+\rho_t\tilde{V}_t(\sigma_{V_t,1}+\lambda_{1,t})dW_{1,t}+\rho_t\tilde{V}_t(\sigma_{V_t,2}+\lambda_{2,t})dW_{2,t}.
\end{aligned}$$

Since $\tilde{V}_t\rho_t$ is a $\mathbb{P}$-martingale, the drift part of above SDE equals to zero and we will get:

$$\mu_{V_t}-V_t r_t+\lambda_{1,t}\sigma_{V_t,1}+\lambda_{2,t}\sigma_{V_t,2}=0.$$

Obviously, the solution for such equation is not unique and $\lambda_{1,t}=-(\mu_{V_t}-r_t)/\sigma_{V_t,1}$ and $\lambda_{2,t}=0$ is one set of answers. Under $\mathbb{Q}$-measure , the SVG model can be rewritten as:

$$dr_t = (\alpha+\beta r_t-\sigma_r r_t^{\gamma}e^{1/2h_t}\lambda_{1,t})dt+\sigma_r r_t^{\gamma}e^{1/2h_t}dW_{1,t}^{\mathbb{Q}} \tag{5.90}$$

$$dh_t = (-\mu h_t-\sigma_h\lambda_{2,t})dt+\sigma_h dW_{2,t}^{\mathbb{Q}} \tag{5.91}$$

where $W_{1,t}^{\mathbb{Q}}$ and $W_{2,t}^{\mathbb{Q}}$ are two independent standard Brownian motions under $\mathbb{Q}$.

5.2.3 SVJ model Under the Risk Neutral Measure

Since the stochastic driven factor in volatility part of the SVJ model has been changed from Brownian motion to the Poisson process, a more general Itô formula is needed when we deal with the SDE of V_t. The Poisson processes in both the SVJ and SVG models are finite activity jump processes, and the corresponding Itô formula is stated by the following proposition:

Proposition 5.1 *(Cont & Tankov (2004), Proposition 8.14) Let X be a diffusion process with jumps, defined as the sum of a drift term, a Brownian stochastic integral and a compound Poisson process:*

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \sum_{i=1}^{N_t} \Delta X_i$$

where b_t and σ_t are continuous nonanticipating processes with

$$E\left(\int_0^T \sigma_t^2 dt\right) < \infty.$$

Then, for any $C^{1,2}$ function $f : [0,T] \times \mathbb{R} \to \mathbb{R}$:

$$df(t, X_t) = \left(\frac{\partial f}{\partial t} + b_t \frac{\partial f}{\partial x} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma_t^2\right) dt + \frac{\partial f}{\partial x}\sigma_t dW_t + \left(f(X_{t-} + \Delta X_t) - f(X_{t-})\right),$$

where $X_{t-} = \lim_{s \to t, s<t} X_s$ and $\Delta X_t = X_t - X_{t-}$.

Given Equation (2.26) and (2.27), the SDE of V_t for the SVJ model is then:

$$dV_t = \mu_{V_t} dt + \sigma_{V_t} dW_t + (V_t(h_{t-} + \Delta h_t) - V_{t-}(h_{t-}))$$

where

$$\begin{aligned} \mu_{V_t} &\equiv \frac{\partial V_t}{\partial t} + \frac{\partial V_t}{\partial r_t}(\alpha + \beta r_t) - \frac{\partial V_t}{\partial h_{t-}}\mu h_{t-} + \frac{1}{2}\frac{\partial^2 V_t}{\partial r_t^2} r_t^{2\gamma} h_{t-}, \\ \sigma_{V_t} &\equiv \frac{\partial V_t}{\partial r_t} r_t^{\gamma} h_{t-}^{1/2}. \end{aligned}$$

For the discounted value $\tilde{V}_t$, the SDE is:

$$d\tilde{V}_t = \mu_{\tilde{V}_t} dt + \sigma_{\tilde{V}_t} dW_t + (\tilde{V}_t(h_{t-} + \Delta h_t) - \tilde{V}_{t-}(h_{t-})), \tag{5.92}$$

where

$$\begin{aligned}\mu_{\tilde{V}_t} &\equiv (\mu_{V_t} - r_t V_t)/B_t, \\ \sigma_{\tilde{V}_t} &\equiv \sigma_{V_t}/B_t.\end{aligned}$$

Therefore $\tilde{V}$ is also a Lévy process with the triplet $(\sigma^2_{\tilde{V}_t}, \nu_{\tilde{V}_t}, \mu_{\tilde{V}_t})$ and with the same Poisson intensity as $\{h_t\}$.

The following two propositions from Cont & Tankov (2004) provide the way to figure out an equivalent martingale measure for a general Lévy process with *characteristic triplet* (σ^2, ν, γ).

Proposition 5.2 *(Cont & Tankov (2004), Proposition 9.8) Let $(X, \mathbb{P})$ and $(X, \mathbb{Q})$ be two Lévy processes on $\mathbb{R}$ with characteristic triplets $(\sigma^2_{\mathbb{P}}, \nu_{\mathbb{P}}, \gamma_{\mathbb{P}})$ and $(\sigma^2_{\mathbb{Q}}, \nu_{\mathbb{Q}}, \gamma_{\mathbb{Q}})$. Then measure $\mathbb{P}$ and $\mathbb{Q}$ given F_t are equivalent for all t if and only if three following conditions are satisfied:*

1. $\sigma^2_{\mathbb{P}} = \sigma^2_{\mathbb{Q}}$.
2. *The Lévy measures are equivalent with*

$$\int_{-\infty}^{\infty} \left(e^{\phi(x)/2} - 1\right)^2 \nu_{\mathbb{P}}(dx) < \infty \tag{5.93}$$

   where $\phi(x) = \ln\left(\frac{d\nu_{\mathbb{Q}}}{d\nu_{\mathbb{P}}}\right)$.

3. *If* $\sigma_{\mathbb{P}} = 0$,

$$\gamma_{\mathbb{Q}} - \gamma_{\mathbb{P}} = \int_{-1}^{1} x(\nu_{\mathbb{Q}} - \nu_{\mathbb{P}})(dx).$$

When $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, the Radon-Nikodým derivative is

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{U_t}$$

with

$$U_t \equiv \eta X^c_t - \frac{\eta^2 \sigma^2_{\mathbb{P}} t}{2} - \eta\gamma_{\mathbb{P}} t + \lim_{\varepsilon \downarrow 0} \left(\sum_{s \le t, |\Delta X_s| > \varepsilon} \phi(\Delta X_s) - t \int_{|x|>\varepsilon} \left(e^{\phi(x)} - 1\right) \nu_{\mathbb{P}}(dx) \right).$$

Here (X_t^c) *is the continuous part of* X_t *and* η *is such that*

$$\gamma_{\mathbb{Q}} - \gamma_{\mathbb{P}} - \int_{-1}^{1} x(\nu_{\mathbb{Q}} - \nu_{\mathbb{P}})(dx) = \sigma_{\mathbb{P}}^2 \eta \tag{5.94}$$

if $\sigma_{\mathbb{P}} > 0$ *and zero if* $\sigma_{\mathbb{P}} = 0$. U_t *is a Lévy process with characteristic triplet* $(\alpha_U, \nu_U, \gamma_U)$ *given by:*

$$\begin{aligned} \alpha_U &\equiv \sigma_{\mathbb{P}}^2 \eta^2 \\ \nu_U &\equiv \nu_{\mathbb{P}} \phi^{-1} \\ \gamma_U &\equiv -\frac{1}{2}\sigma_{\mathbb{P}}^2 \eta^2 - \int_{-\infty}^{\infty} (e^y - 1 - y 1_{|y|\leq 1})(\nu_{\mathbb{P}} \phi^{-1})(dy). \end{aligned}$$

Proposition 5.3 *(Cont & Tankov (2004), Proposition 3.18) Let* $(X_t)_{t\geq 0}$ *be a Lévy process on* $\mathbb{R}$ *with characteristic triplet* (σ^2, ν, γ). (X_t) *is a martingale if and only if*

$$\int_{|x|\geq 1} x\nu(dx) < \infty,$$

and

$$\gamma + \int_{|x|\geq 1} x\nu(dx) = 0.$$

Then under the equivalent martingale measure $\mathbb{Q}$, the drift of $\tilde{V}_t$ will satisfy:

$$\mu_{\tilde{V}_t}^{\mathbb{Q}} = -\int_{|x|\geq 1} x\nu_{\tilde{V}_t}^{\mathbb{Q}}(dx),$$

where $\nu_{\tilde{V}_t}^{\mathbb{Q}}(dx) = e^{\phi(x)}\nu_{\tilde{V}_t}(dx)$ with $\phi(x)$ satisfying the condition (5.93). The volatility remains unchanged: $\sigma_{\tilde{V}_t} = \sigma_{\tilde{V}_t}^{\mathbb{Q}}$. Therefore with different choices of $\phi(x)$ we will obtain various kinds of equivalent martingale measure. Same as the SVG model, the market is incomplete when the underlying process follows the SVJ model. For example, let $\phi(x) = 0$ which indicates under the equivalent measure $\mathbb{Q}$, the new Lévy measure $\nu_{\tilde{V}_t}^{\mathbb{Q}}$ is unchanged from $\mathbb{P}$-measure. By Proposition 5.3, $\mu_{\tilde{V}_t}^{\mathbb{Q}} = -\int_{|x|\geq 1} x\nu_{\tilde{V}_t}(dx)$. The η in Proposition 5.2 will then have the following specification:

$$\eta_t = -\frac{\int_{|x|\geq 1} x\nu_{\tilde{V}_t}(dx) + \mu_{\tilde{V}_t}}{\sigma_{\tilde{V}_t}^2}. \tag{5.95}$$

As a result, we find a *risk-netrual* measure $\mathbb{Q}$ with the Radon-Nikodým derivative:

$$\rho_t = \exp\left\{\int_0^t \eta_s \sigma_{\tilde{V}_s} dW_s - \frac{1}{2}\int_0^t \eta_s^2 \sigma_{\tilde{V}_s}^2 dt\right\}. \tag{5.96}$$

Under such $\mathbb{Q}$-measure, the SDE of r_t will have the following specification:

$$\begin{aligned} dr_t &= (\alpha - \beta r_t - r_t^{2\gamma} h_t \eta_t)dt + r_t^{\gamma} h_t^{1/2} dW_t^{\mathbb{Q}}, \\ dh_t &= -\mu h_t dt + dX_t, \end{aligned}$$

where $W_t^{\mathbb{Q}}$ is a standard Brownian motion under $\mathbb{Q}$-measure, and with the same compound Poisson process h_t.

5.2.4 SVS model Under the Risk Neutral Measure

For the SVS model, V_t is a function of (t, r_t, h_t) where $h_t = h_{1.t} + h_{2,t}$. Applying the general Itô formula in Proposition 5.1, we obtain the SDE of V_t:

$$dV_t = \mu_{V_t} dt + \sigma_{V_t} dW_t + (V_t(h_{t-} + \Delta h_t) - V_t(h_{t-})) \tag{5.97}$$

where

$$\begin{aligned} \mu_{V_t} &\equiv \frac{\partial V_t}{\partial t} + \frac{\partial V_t}{\partial r_t}(\alpha + \beta r_t) - \frac{\partial V_t}{\partial h_{1,t-}}\mu_1 h_{1,t-} - \frac{\partial V_t}{\partial h_{2,t-}}\mu_2 h_{2,t-} + \frac{1}{2}\frac{\partial^2 V_t}{\partial r_t^2} r_t^{2\gamma} h_{t-} \\ \sigma_{V_t} &\equiv \frac{\partial V_t}{\partial r_t} r_t^{\gamma} h_{t-}^{0.5}. \end{aligned}$$

For the discounted value $\tilde{V}_t$, the SDE is:

$$d\tilde{V}_t = \mu_{\tilde{V}_t} dt + \sigma_{\tilde{V}_t} dW_t + (\tilde{V}_t(h_{t-} + \Delta h_t) - \tilde{V}_t(h_{t-})) \tag{5.98}$$

where

$$\begin{aligned} \mu_{\tilde{V}_t} &\equiv B_t^{-1}(\mu_{\tilde{V}_t} - r_t V_t) \\ \sigma_{\tilde{V}_t} &\equiv B_t^{-1}\sigma_{\tilde{V}_t}. \end{aligned}$$

Therefore, $\tilde{V}_t$ is a Lévy process with characteristic triplet $(\sigma_{\tilde{V}_t}^2, \nu_{\tilde{V}_t}, \mu_{\tilde{V}_t})$ where $\nu_{\tilde{V}_t}$ is the Lévy measure for $\tilde{V}_t$ which has the same Poisson intensity as the volatility driving factor h_t in the SVS model.

As the SVJ model, we set $\phi(x) = 0$, and according to the Proposition 5.2 and 5.3, the *risk-neutral* measure $\mathbb{Q}$ for the SVS model can be found and the Radon-Nikodým derivative has the following specification:

$$\rho_t = \exp\left\{\int_0^t \eta_s \sigma_{\tilde{V}_s} dW_s - \frac{1}{2}\int_0^t \eta_s^2 \sigma_{\tilde{V}_s}^2 dt\right\},$$

where

$$\eta_t = -\frac{\int_{|x|\geq 1} x\nu_{\tilde{V}_t}(dx) + \mu_{\tilde{V}_t}}{\sigma_{\tilde{V}_t}^2}.$$

Under $\mathbb{Q}$-measure, the SVS model will have specification like:

$$\begin{aligned} dr_t &= (\alpha + \beta r_t - r_t^{2\gamma} h_t \eta_t)dt + r_t^{\gamma} h_t^{1/2} dW_t^{\mathbb{Q}}, \\ h_t &= h_{1,t} + h_{2,}, \\ dh_{i,t} &= -\mu_i h_{i,t} dt + dX_{i,t}, i = 1, 2, \end{aligned}$$

where $W_t^{\mathbb{Q}}$ is a standard Brownian motion under $\mathbb{Q}$-measure, and with the same compound Poisson processes $h_{i,t}, i = 1, 2$ under $\mathbb{P}$-measure.

5.3 Pricing Zero Coupon Bond

Under the *risk-neutral* measure $\mathbb{Q}$, the price of a zero-coupon bond at time t with one unit final payment at maturity T, $P(t,T)$, is the expectation of the discounted value of its final payment:

$$P(t,T) = E^{\mathbb{Q}}\left(e^{-\int_t^T r_s ds}|\mathcal{F}_t\right). \tag{5.99}$$

Recall that the yield to maturity of the T-bond is the spot rate defined as:

$$R(t,T) = -\frac{\log P(t,T)}{T-t},$$

and such spot rates with different maturities construct the yield curve which is an important tool for the studies of both economics and finance. The yield curve is usually used as a benchmark for other debt rates and to help the prediction of future movements of the market. The shape of yield curve is various. The most observed

yield curve is called *normal curve* in which the yields rise as the maturities increasing and the slope is positive but decreasing. When the long term yield is much higher than the short yield, the yield curve is called *steep curve*. An *inverted curve* arises when the short yield is higher than the long one. *Humped curves* are also observed in market which are the curves with almost same short yields and long yields but the middle term ones are either above or below them.

Therefore, yield curve changes are not merely the parallel shifts. Studies of Litterman & Scheinkman (1991) identify three common factors which can explain most of the variations of yield curve, named as *level*, *steepness*, and *curvature*. Generally, the *level* factor represents parallel shifts in yield curves indicating the effect will be the same for rates of all maturities; the affection of the *steepness* factor will be more significant for yields of short maturities than long term one, then it will change the slope of the curve (it is also called as *slope* factor); the affection of the third factor, *curvature*, works mainly on the medium term yields which will make hump-shaped curves. A good candidate of interest rate model is not only able to generate many types of yield curve but also to bring similar changes of them. As we will show in the following sections, yield curves generated by SV models are much richer than the CKLS model and jump models perform better than SVG model at capturing these three dominant factors.

Since analytical pricing formulas cannot be found for these models, we will use Monte Carlo method to value fixed-income products. Using Monte Carlo method, the estimated value of the price, $P(t,T)$ from Equation (5.99) is calculated as:

$$\hat{P}(t,T) = \frac{1}{M}\sum_{j=1}^{M} e^{-\sum_{i=1}^{N} r_{i,j}\Delta t} \tag{5.100}$$

where N is the length of sample path with $N = \frac{T-t}{\Delta t}$, Δt is one week, M is the number of parallel sample trials, $t_0 = t, t_k = t_0 + \Delta t \times k, t_N = T$ and $r_{i,j}$, is the jth sample path between t_{i-1} and t_i for $j = 1, ..., M$ which is constructed from its discrete process. In this and following sections, we will set the market price of risk equal to zero, therefore the risk neutral measure $\mathbb{Q}$ will identical to the canonical

measure $\mathbb{P}$. Under the risk-neutral measure, the Euler discretization of short rate process has the following form:

$$r_{i+1,j} = r_{i,j} + (\alpha + \beta r_{i,j})\Delta t + \sigma(r_{i,j}, h_{i,j})\varepsilon_{i+1,j}$$

where $\{\varepsilon_{i+1,j}\}_{i=t}^{T-1}$ is a series of $i.i.d$ random variables with normal distribution $N(0, \Delta t)$ under $\mathbb{Q}$. We will generate the short rate processes from various Euler discretizations of candidate models in following sections. The parameters used in these models are the posterior means listed in Table 4.7 to Table 4.11 respectively which are the results of MCMC procedure described in Chapter 3. The total number of sample paths for each model is 1 million.

5.3.1 Yield curve from CKLS model

The jth simulation of short rate process for the CKLS model is generated from:

$$r_{i+1,j} = r_{i,j} + (\alpha + \beta r_{i,j})\Delta t + \sigma_r r_{i,j}^{\gamma}\varepsilon_{i+1,j}$$

where $\varepsilon_{i+1,j} \sim N(0, \Delta t)$ and the values of parameters are the estimations from Chapter 4. Here for the CKLS model and in the following sections for the SV models, we will improve the Monte Carlo simulation by applying a speed-up method: *the antithetic variate method* (see examples in James & Webber (2000)). Under this method, two sample paths are obtained from one set of random variables $\varepsilon_{i,j}$ by changing $\varepsilon_{i,j}$ to $-\varepsilon_{i,j}$. The sample variance will be much more reduced by applying this method. For example, the variance of 2-year zero coupon bond is only around 10% of the one from normal Monte Carlo simulation and the variance of 10-year bond reduced also to almost 20% of the normal one.

For the single factor CKLS model, we investigate the various shapes of yield curve by changing $r_{0,.}$. In Figure 5.41, we can observe the *normal curve* when r_0 is below the long term mean $-\alpha/\beta = 0.0637$; *humped curve* can be found when r_0 is around the mean; and when r_0 is above the mean, we obtain the *inverted curve*. However other types of curves can not be obtained from CKLS model by merely changing

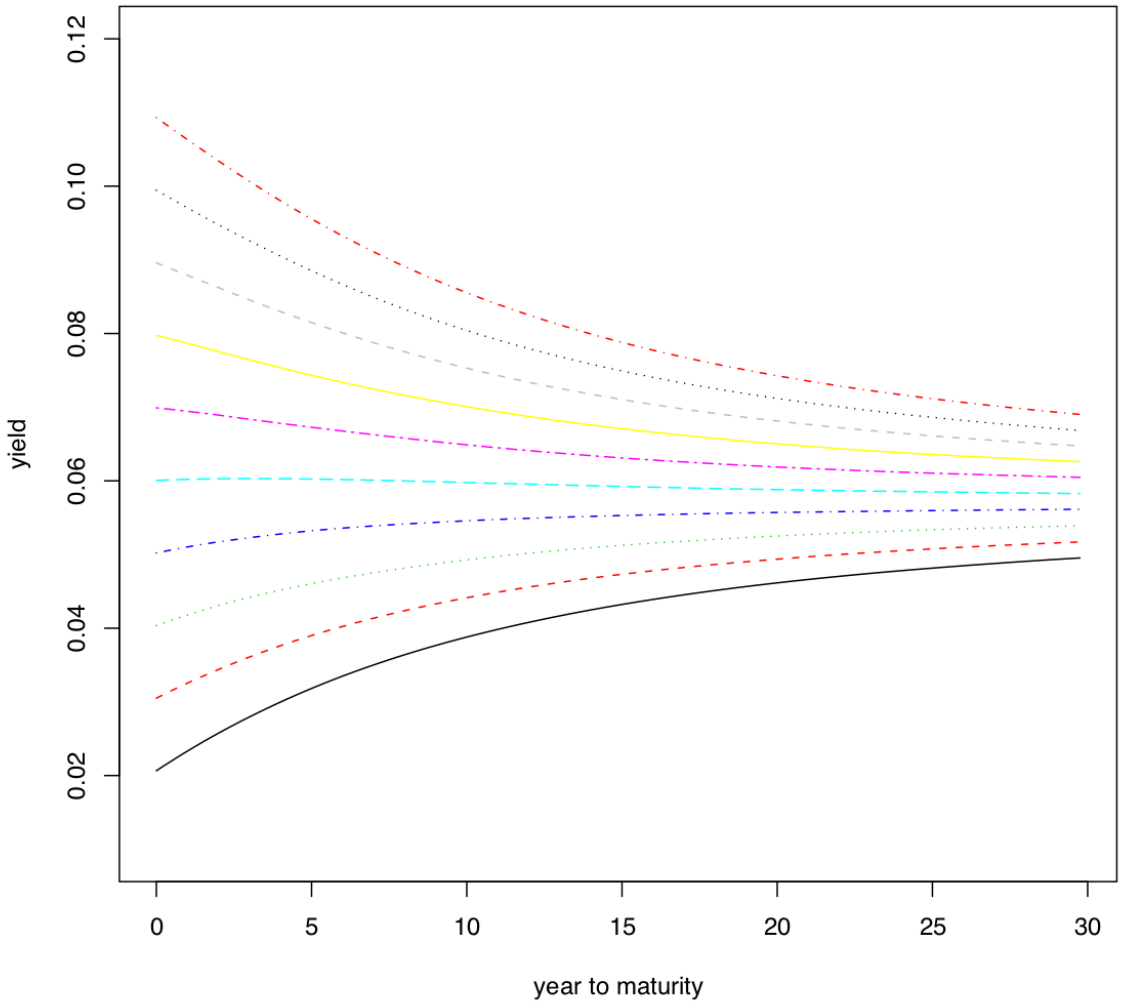

Figure 5.41: Sample yield curves for the CKLS model: r_0 started from 0.02 to 0.11 with increment 100bps (0.01).

the starting value of r_t, like *inverted hump* for example.

While r_0 changes from one point to other nearest points, the whole yield curve generally moves in parallel which is the effect of *level factor*. The change of steepness of the curve can be observed when r_0 crosses over the long term mean although the yield curve does not converge quickly to the mean. We cannot observe the effect of third factor, *curvature*, from the graph due to the property of single factor model that there is only one variable available to be changed.

5.3.2 Yield curve from SVG model

The jth simulation of short rate process for the SVG model is generated from:

$$\begin{aligned} r_{i+1,j} &= r_{i,j} + (\alpha + \beta r_{i,j})\Delta t + \sigma_r e^{h_{i,j}/2} r_{i,j}^{\gamma} \varepsilon_{i+1,j} \\ h_{i+1,j} &= h_{i,j} - \mu h_{i,j} \Delta t + \sigma_h \xi_{i+1,j} \end{aligned}$$

where $\varepsilon_{i+1,j} \sim N(0, \Delta t)$ and $\xi_{i+1,j} \sim N(0, \Delta t)$ are two independent random variables under $\mathbb{Q}$.

Figure 5.42 plots the yield curves with variant r_0 and σ_0 and the initial volatility for the SVG model is defined by $\sigma_0 = \sigma_r e^{h_{0,j}/2} r_{0,j}^{\gamma}$. The top plot presents the changing of r_0 from 0.01 to 0.11 under a smaller starting volatility where $\sigma_0 = 0.01$. The bottom one shows the changing with same range but a higher value of σ_0 where $\sigma_0 = 0.1$. We can observe *normal curve*, *humped curve* and *inverted curve* from both plots as from the CKLS model. Similarly as CKLS model, these plots provide the evidence of both *level* effect and *steepness* effect by changing r_0 in SVG model. However, attention should be paid that, above the long term mean of SVG model which equals to 0.0633, the curvature of yield curves increases as σ_0 changes from 0.01 to 0.1.

Figure 5.43 shows how the yield curve depends on σ_0. We present such changes given different r_0: below the long term mean, top plot; around the mean, medium one and

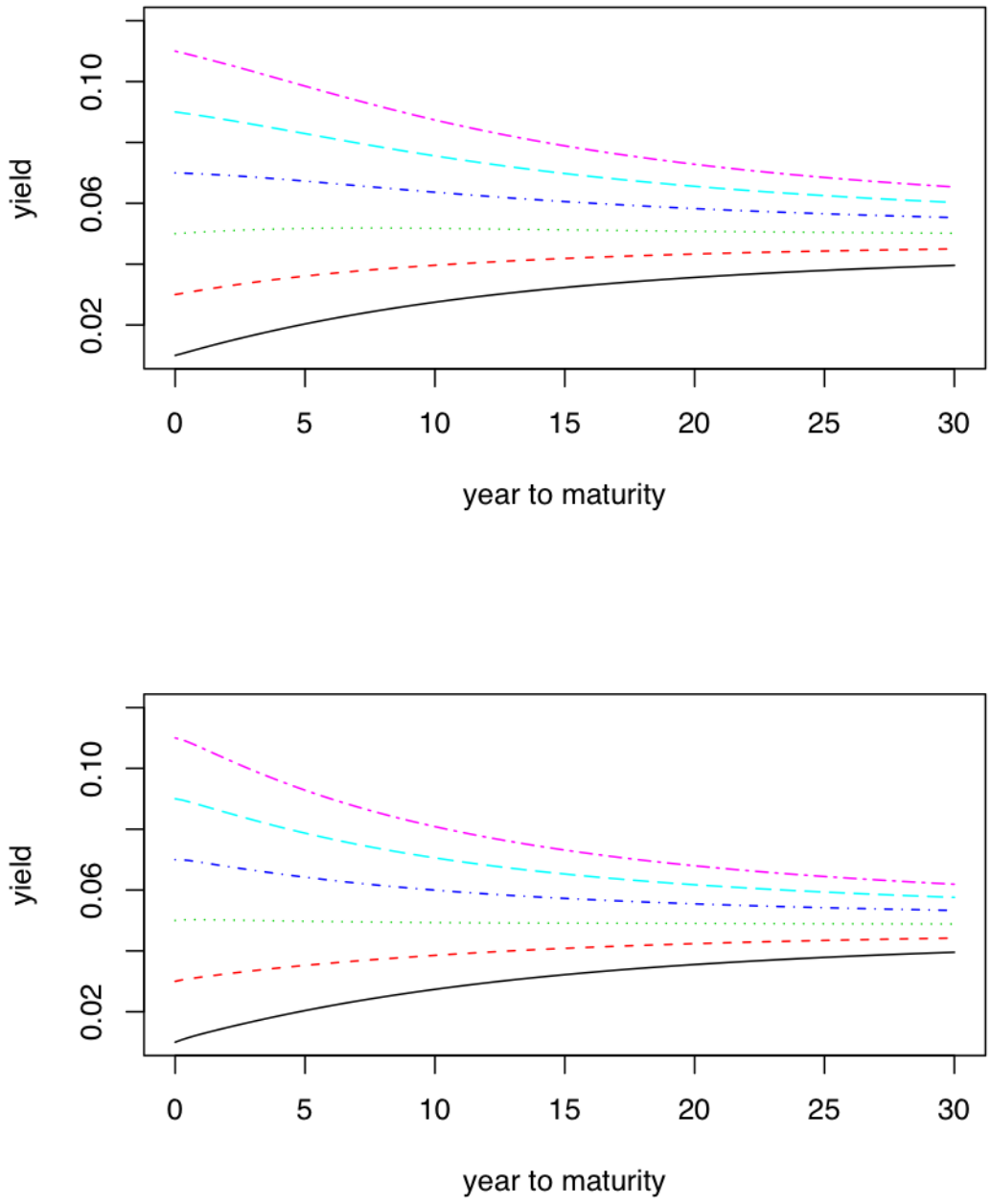

Figure 5.42: Sample yield curves for the SVG model changing with r_0 from 0.01 to 0.11 with increment 200bps (0.02): Top: with $\sigma_0 = 0.01$; Bottom: with $\sigma_0 = 0.1$.

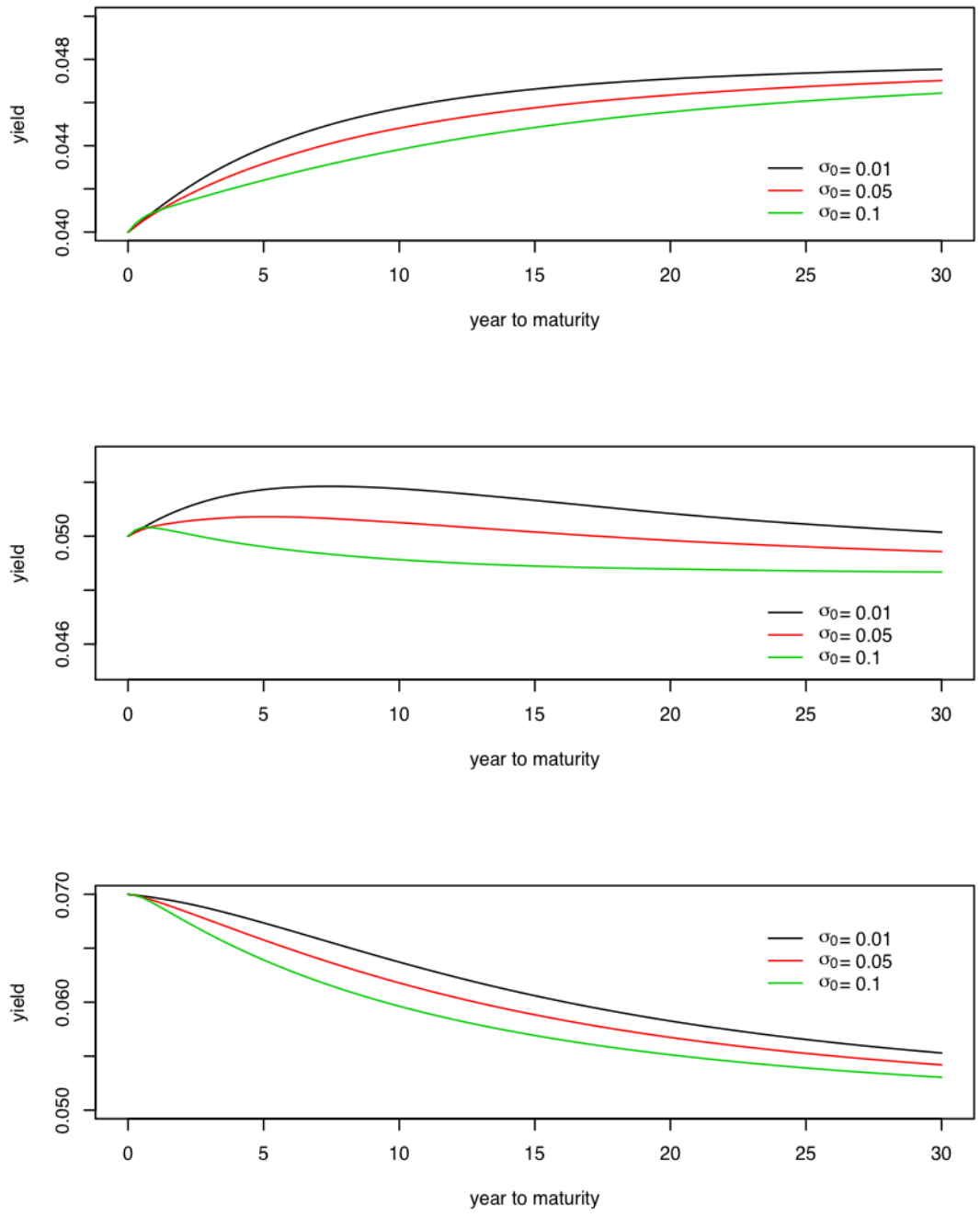

Figure 5.43: Sample yield curves for the SVG mode changing with σ_0. Top: $r_0 = 0.04$. Middle: $r_0 = 0.05$. Bottom: $r_0 = 0.07$.

above the mean, bottom one. Obviously, the increase of σ_0 causes the decrease of yield, and the effect is bigger on the period of 5-year to 15-year than others. This coincides with the results of Litterman & Scheinkman (1991) that the changes in curvature of the yield curve are associated with the changes in rate volatility.

5.3.3 Yield curve from SVJ model

Under the $\mathbb{Q}$-measure, the jth simulation of short rate process for the SVJ model is generated from:

$$\begin{aligned} r_{i+1,j} &= r_{i,j} + (\alpha + \beta r_{i,j})\Delta t + h_{i,j}^{1/2} r_{i,j}^{\gamma} \varepsilon_{i+1,j} \\ h_{i+1,j} &= h_{i,j} - \mu h_{i,j}\Delta t + \sigma_h \Delta X_{i,j} \\ \Delta X_{i,j} &= \sum_{s=1}^{N_{i,j}} \xi_{s,j} \end{aligned}$$

where $\varepsilon_{i+1,j} \sim N(0, \Delta t)$, $N_{i,j}$ is the number of jumps in period $(i, i+\Delta t)$, $\xi_{s,j}$ is the jump size generated from standard exponential distribution in our case.

Figure 5.44 presents the yield curves with various r_0 from 0.01 to 0.11 given different initial volatility σ_0 which is determined by $\sigma_0 = h_{0,j}^{1/2} r_{0,j}^{\gamma}$. The top plot shows such curves given a lower initial value of volatility whereas those curves with bigger σ_0 are presented in the bottom plot. Here we observe similar yield curve shapes to the CKLS and SVG models. The *steepness* effect is more significant in the bottom plot where the curves simulated with higher σ_0. Comparing the bottom graph in Figure 5.44 with the bottom one in Figure 5.42, the curvature of the SVJ yield curves is much bigger than the SVG's given the same initial value of both r_0 and σ_0.

The effect on the yield curves as σ_0 changes are presented in Figure 5.45 for three levels of r_0. Comparing with Figure 5.43, the *curvature* effect is more significant for the SVJ yield curves than SVG's. As σ_0 moves to 0.1, the short-term yields is increasing first and then decreasing sharply which make a S-shaped yield curve which is a feature that we did not see with the CKLS and SVG models. Two factors caused the S-shape yield curve for the SVJ model. Firstly the effect of higher initial

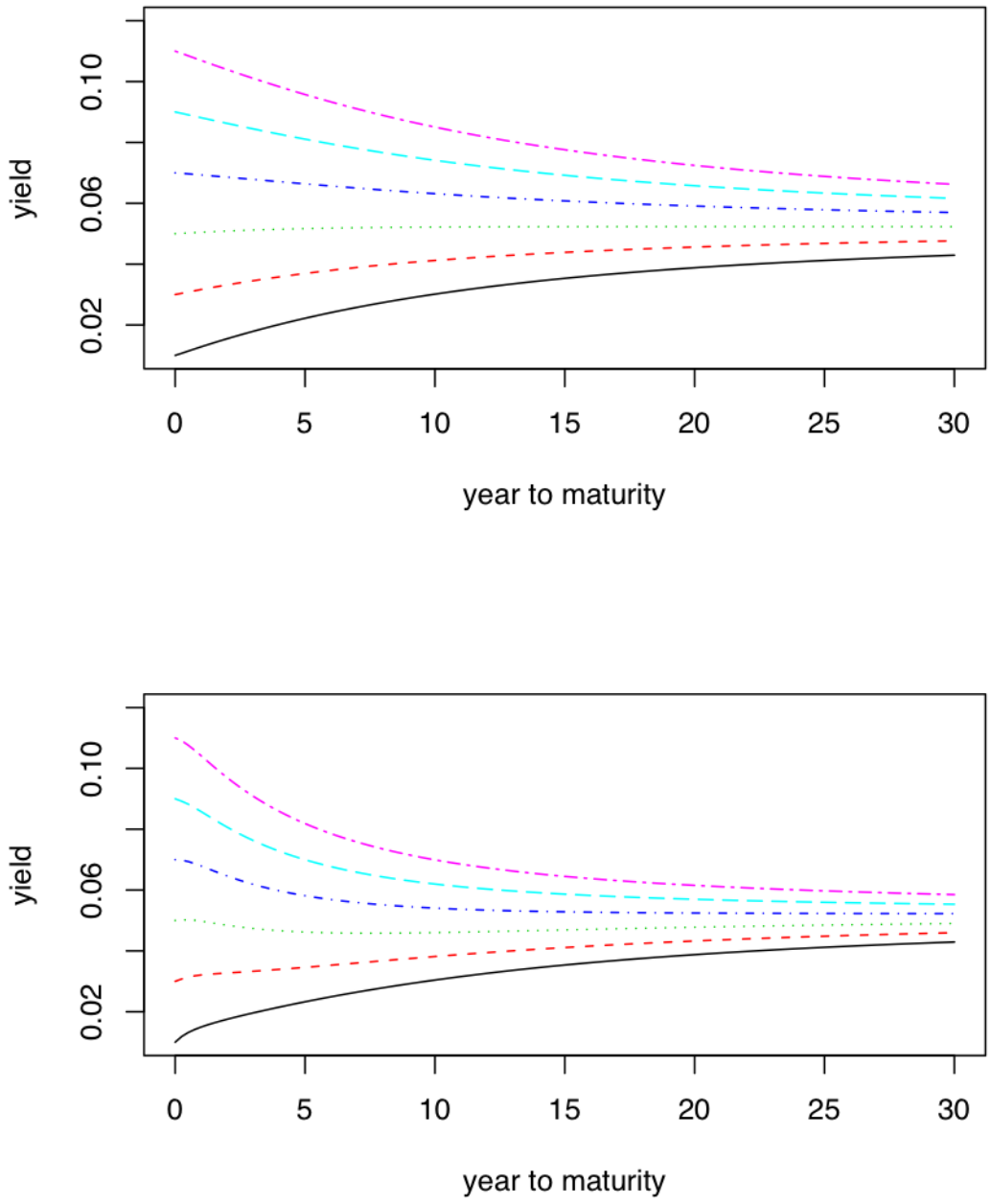

Figure 5.44: Sample yield curves for the SVJ model changing with r_0 from 0.01 to 0.11 with increment 200bps (0.02): Top: with $\sigma_0 = 0.01$; Bottom: with $\sigma_0 = 0.1$.

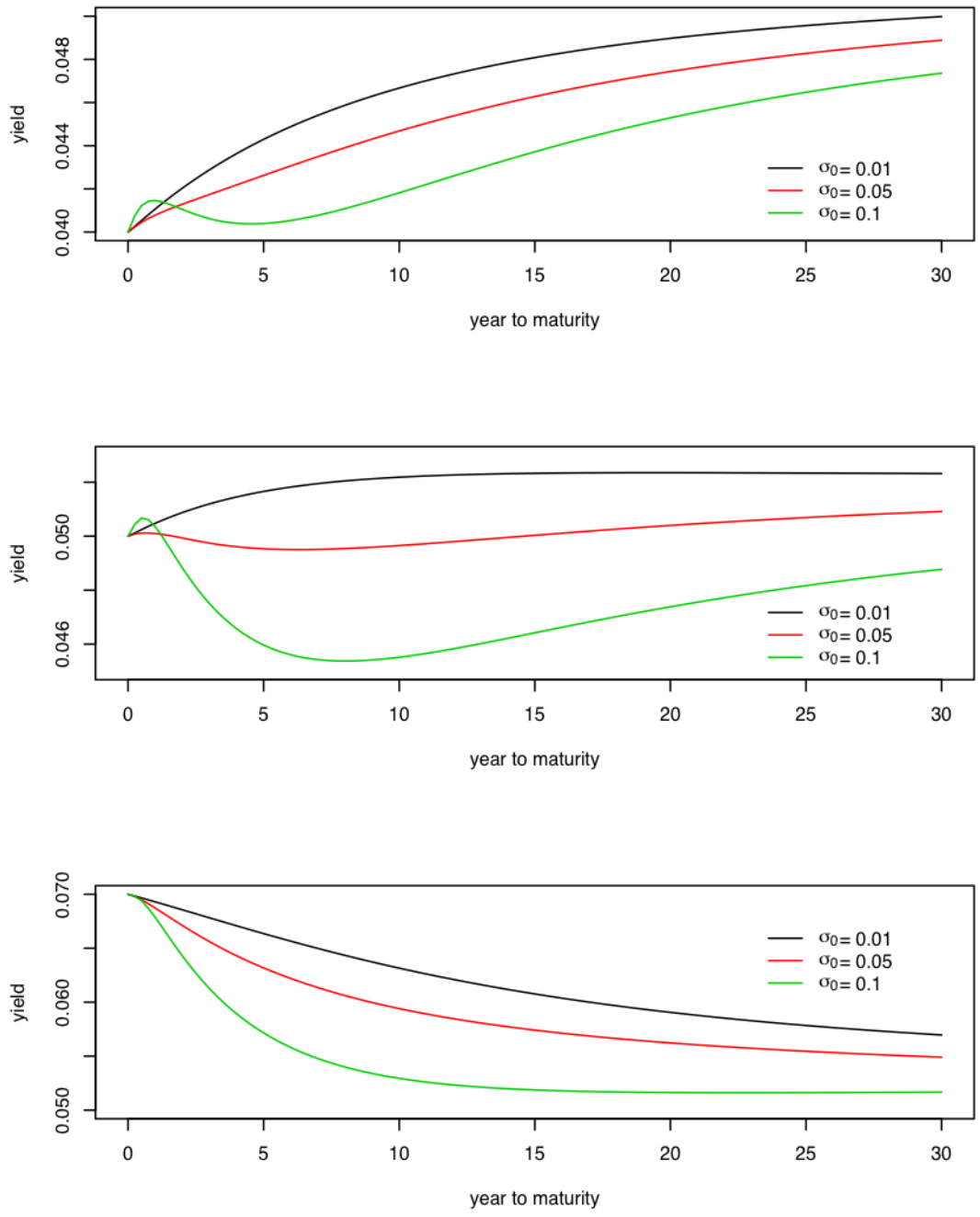

Figure 5.45: Sample yield curves for the SVJ model changing with σ_0. Top: $r_0 = 0.04$. Middle: $r_0 = 0.05$. Bottom: $r_0 = 0.07$.

volatility is more significant when the initial rate is low which can be observed not only from the SVJ model but also from Figure 5.43 for the SVG model and Figure 5.47 for the SVS model. Secondly, since the volatility decays exponentially in the SVJ model and it will take longer time for the effect to decrease than the other two SV models.

5.3.4 Yield curve from SVS model

The jth simulation of short rate process for the SVS model is generated from:

$$\begin{aligned} r_{i+1,j} &= r_{i,j} + (\alpha + \beta r_{i,j})\Delta t + h_{i,j}^{0.5} r_{i,j}^{\gamma} \varepsilon_{i+1,j} \\ h_{i+1,j} &= \sum_{k=1}^{2} h_{i+1,j,k} \\ h_{i+1,j,k} &= h_{i,j,k} - \mu_k h_{i,j,k}\Delta t + \sigma_h \Delta X_{i,j,k}, \quad \forall k = 1,2 \\ \Delta X_{i,j,k} &= \sum_{s=1}^{N_{i,k}} \xi_{s,j,k}, \quad \forall k = 1,2 \end{aligned}$$

where $\varepsilon_{i+\Delta t,j} \sim N(0, \Delta t)$ under $\mathbb{Q}$-measure, $N_{i,j,k}$ is the number of jumps in period $(i, i+\Delta t)$ of the kth OU process , $\xi_{s,j,k}$ is the kth OU jump process size from standard exponential distribution.

The initial volatility σ_0 we used equals to $r_0^{\gamma}\sqrt{h_{0,1} + h_{0,2}}$ with $h_{0,1} = h_{0,2}$. The yield curves simulated from the SVS model are presented in Figure 5.46 and Figure 5.47. The *level* effect can be observed when r_0 moves in small increments in both graphs in Figure 5.46. The yield curve is much steeper as r_0 moves to higher level. This can be observed more clearly when the initial volatility is higher as shown in the bottom graph. The curvature of SVS's yield curves is bigger than the SVG's but smaller than those from the SVJ model which shows the effect of the volatility jumps in bond value. As we have shown in section 4.9.4, the volatility process from the SVS model evolves mostly under the process from SVJ model. Then the effect of volatility is less than the SVJ model.

The *curvature* effect can also be observed from all three graphs in Figure 5.47. The

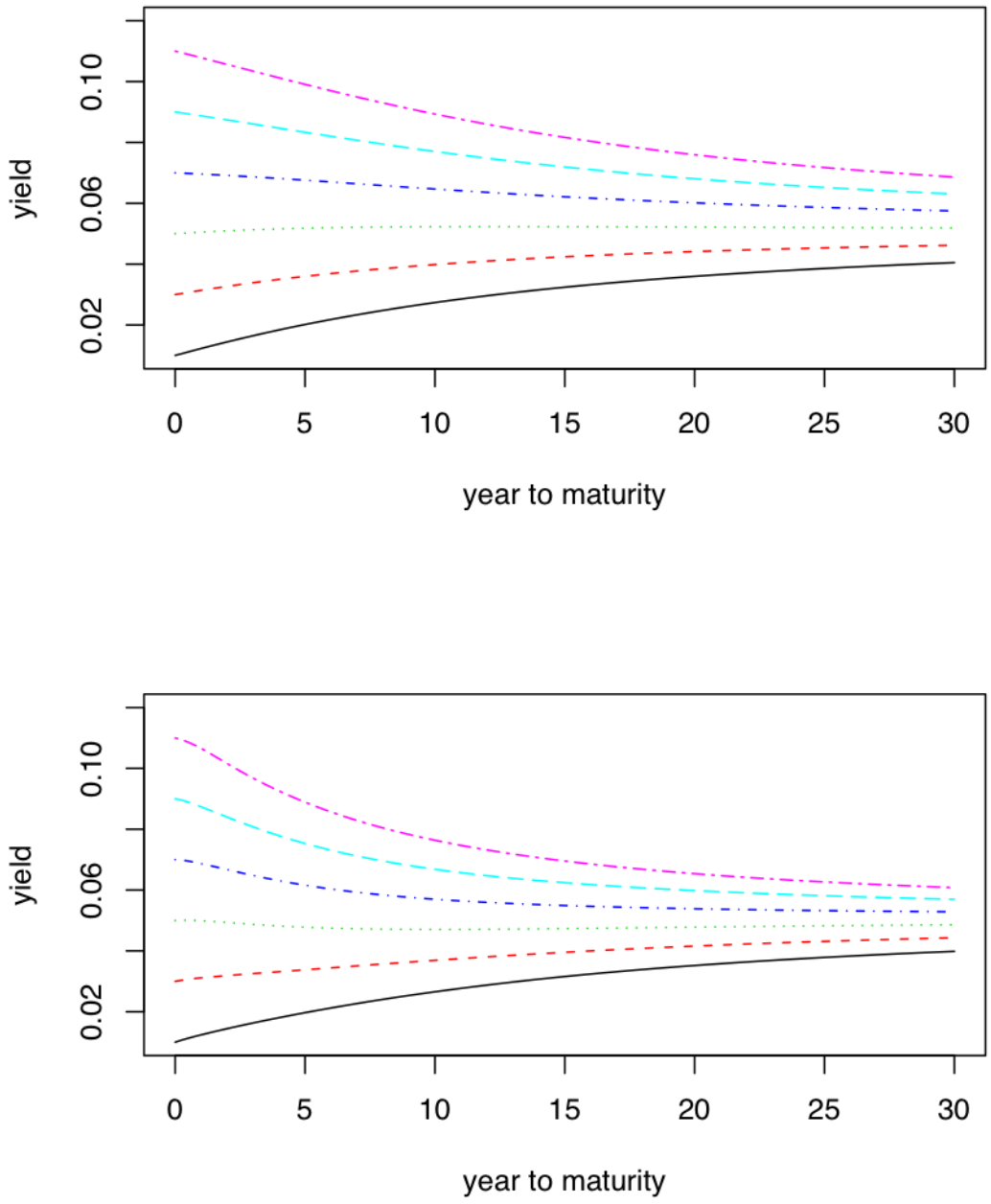

Figure 5.46: Sample yield curves for the SVS model changing with r_0 from 0.01 to 0.11 with increment 200bps (0.02): Top: with $\sigma_0 = 0.01$; Bottom: with $\sigma_0 = 0.1$.

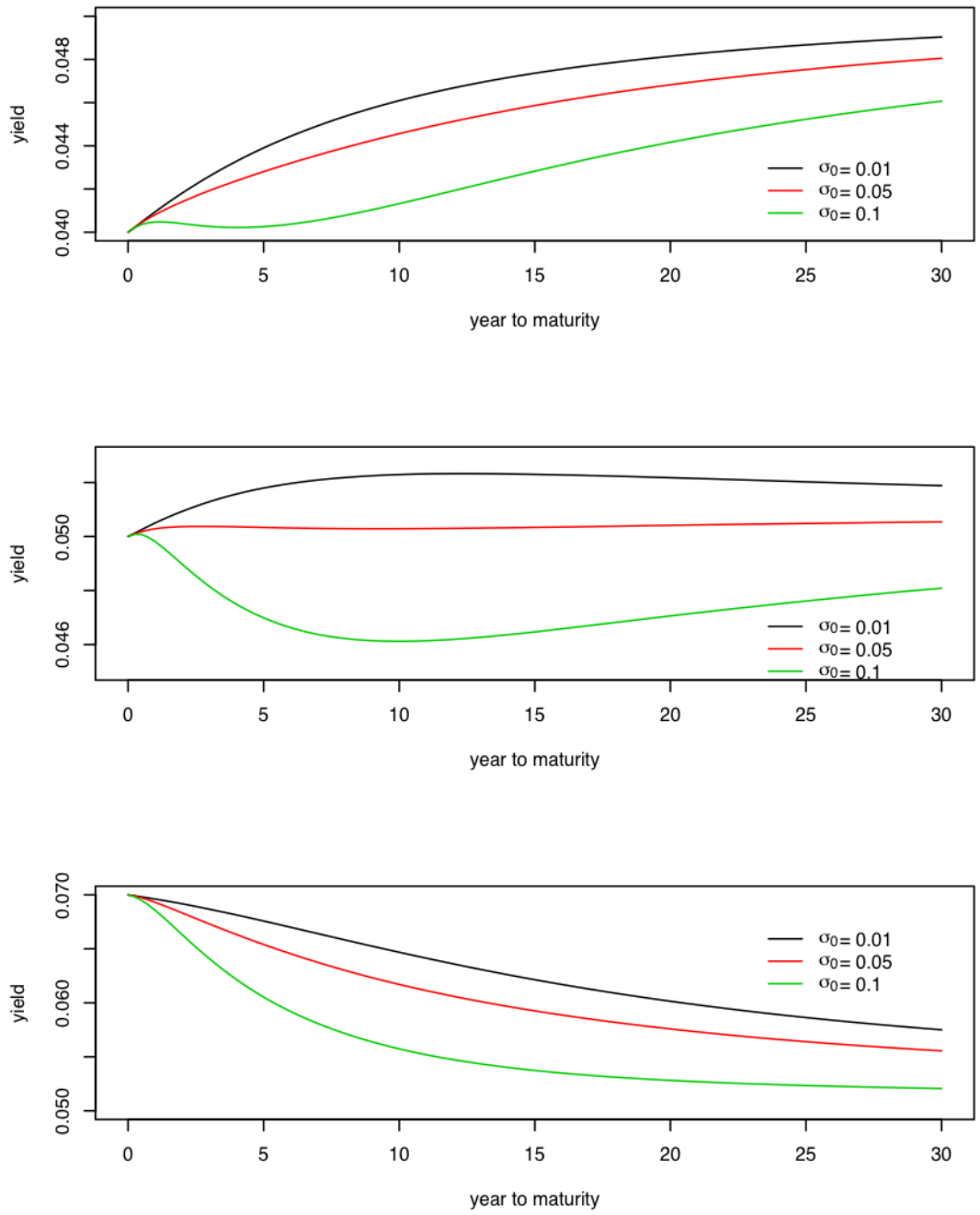

Figure 5.47: Sample yield curves for the SVS model changing with σ_0. Top: $r_0 = 0.04$. Middle: $r_0 = 0.05$. Bottom: $r_0 = 0.07$.

similar S-shape yield curve can be found from the top graph but less significant than those from the SVJ model due to more consistent smaller jump process added in the SVS model. Shown in the middle graph, the yield curve has a *inverted-hump* shape when σ_0 moves to 0.1 which is more significant in the jump models comparing with the SVG model.

5.4 Pricing Zero Coupon Bond Option

5.4.1 Valuation of European call option

Under the *risk-neutral* measure $\mathbb{Q}$, the process of discounted asset value, $\{\tilde{H}_T =: \tilde{H}_T = e^{-\int_0^T r_t dt} H_T\}$, is a martingale. The price of such derivative at t, H_t, is determined by:

$$H_t = E^{\mathbb{Q}} \left(e^{-\int_t^T r_s ds} H_T | \mathcal{F}_t \right). \tag{5.101}$$

If $H_T = 1$, H_t is the value of a zero coupon bond at time t as we have discussed in previous sections.

In this section, we will consider the European Call option on a zero coupon bond, $P(0,S)$, with maturity S. The payoff at expiring time T ($T < S$), $H_T = \max(P(T,S) - K, 0)$ where K is the strike price. According to Equation (5.101), the arbitrage-free price of such option is given by:

$$c_0(T,S) = E^{\mathbb{Q}} \left(e^{-\int_0^T r_t dt} \max(P(T,S) - K, 0) | \mathcal{F}_0 \right). \tag{5.102}$$

For a put option with same conditions, the price is:

$$p_0(T,S) = E^{\mathbb{Q}} \left(e^{-\int_0^T r_t dt} \max(K - P(T,S), 0) | \mathcal{F}_0 \right), \tag{5.103}$$

which can also be obtained from the *put-call parity*:

$$p_0(T,S) = c_0(T,S) + KP(0,T) - P(0,S). \tag{5.104}$$

Since there are no analytical formulas for option pricing based on the short rate models under consideration, Monte Carlo simulation and the antithetic variate method described in Section 5.3 will be applied. Then the simulated value of call option defined by Equation 5.102 can be obtained through the following equation:

$$\hat{c}_0(T,S) = \frac{1}{M}\sum_{j=1}^{M} e^{-\sum_{i=1}^{N} r_{i,j}\Delta t} \max(\hat{P}_j(T,S) - K, 0),$$

where M is the total number of Monte Carlo sample paths, $N = T/\Delta t$ with Δt for one week, and $\hat{P}_j(T,S)$ is the estimated bond price for the jth sample path at the option expiring date T. $\hat{P}_j(T,S)$ can be obtained by simply running further Monte Carlo simulations after T to S which means:

$$\hat{P}_j(T,S) = \exp\left\{-\sum^{N'} r_{i,j}\Delta t\right\}, \tag{5.105}$$

here the jth sample path of $r_{i,j}$ starting from expiration T ending at maturity S and the length of this path $N' = (S-T)/\Delta t$. Obviously, it is a very rough way to estimate bond value through just single sample path. The simulation variance can be reduced if we general M' parallel sample paths from time T to S, and the value will be:

$$\hat{P}_j(T,S) = \frac{1}{M'}\sum_{k=1}^{M'} \exp\left\{-\sum^{N'} r_{i,j,k}\Delta t\right\}, \tag{5.106}$$

where $r_{i,j,k}$ represents the kth sample of short rate between (T,S) starting from the jth run for discount factor. This is a time-consuming method which requiring $M \times M'$ sample paths. Is there any other efficient method can be applied for $\hat{P}_j(T,S)$? The answer is "Yes", as we will see, it is the Least-squares Monte Carlo approach.

5.4.2 Least-squares Monte Carlo approach

The least-squares Monte Carlo approach introduced by Longstaff & Schwartz (2001) is used here to value $\tilde{P}_j(T,S)$. This methodology, in short as LSM offers a reasonable and efficient approach to simulate the value of underlying in option pricing.

the algorithm

The LSM approach uses least squares to approximate the conditional expectation function (5.101) at T using an analytical function. We assume that the unknown functional form of $P(T,S)$ can be represented as a linear combination of finite set of F_T measurable basis functions:

$$\hat{P}(T,S) = \sum_{i=1}^{M} a_i f_i(X), \tag{5.107}$$

where M is the number of total basis functions, a_1, ..., a_M are the constant coefficients, X is the state vector.

Various candidate forms can be considered as basis functions, such as the Laguerre polynomials, Fourier series, simple powers of the state variables and so on. Longstaff & Schwartz (2001) point out that all these candidates can give accurate results. For higher dimensional problems, with two state variables X and Y for example, the set of basis functions should include terms in both X and Y, as well as cross-products of these terms. Here we will use the first two powers of state variables as the base function, and the approximating function will be:

$$F(r,\sigma) = a_0 + a_1 r_T + a_2 r_T^2 + a_3 \sigma_T + a_4 \sigma_T^2 + a_5 r_T \sigma_T, \tag{5.108}$$

where r_T is the short rate at T and σ_T is defined differently for each model. For the CKLS model, $\sigma_T = 0$, $\sigma_T = e^{h_T/2}$ for the SVG model, $\sigma_T = h_T^{1/2}$, and for the SVS model, $\sigma_T = h_T^{1/2}$ where $h_T = h_{1,T} + h_{2,T}$.

We start with the naive estimation mentioned in section 5.4.1. The estimation of $P(T,S)$ for the jth Monte Carlo run is given by:

$$\tilde{P}_j(T,S) = \exp\left(-\sum_{i=1}^{N} r_{i,j}\Delta t\right). \tag{5.109}$$

where $r_{1,j}$ is the short rate at time T for the jth simulation and $r_{N,j}$ is the rate at time S, and $N = (S-T)/\Delta t$. With the recorded value of both r and σ at time T for each simulation, we can find the value of parameters, $\{a_k\}$ of $F()$ defined in

Table 5.16: The ratio of 1-year zero-bond price from LSM to Monte Carlo Simulation

	r_0	0.01	0.03	0.05	0.07	0.10
CKLS		1.000010	1.000010	0.999999	1.000000	0.999994
	σ_0	0.01	0.03	0.05	0.07	0.10
SVG	0.0005	0.999992	0.999977	1.000002	0.999971	1.000002
	0.0010	1.000010	0.999983	1.000010	1.000010	1.000030
	0.0020	1.000010	0.999993	0.999973	1.000040	1.000000
	0.0030	1.000010	0.999982	0.999980	0.999988	0.999965
	0.0040	0.999987	1.000000	1.000020	0.999992	0.999983
	σ_0	0.01	0.03	0.05	0.07	0.10
SVJ	0.0005	0.999993	1.000040	0.999975	0.999979	1.000010
	0.0010	1.000010	0.999958	0.999996	1.000020	0.999904
	0.0020	0.999950	0.999974	1.000050	1.000040	1.000010
	0.0030	0.999970	0.999963	0.999976	1.000060	0.999965
	0.0040	0.999886	0.999829	1.000170	1.000060	1.000110
	σ_0	0.01	0.03	0.05	0.07	0.10
SVS	0.0005	1.000010	0.999974	1.000030	0.999961	1.000030
	0.0010	0.999999	1.000030	0.999960	0.999992	1.000010
	0.0020	0.999991	1.000030	1.000070	1.000230	1.000200
	0.0030	0.999966	1.000120	0.999988	0.999924	0.999909
	0.0040	0.999941	0.999991	1.000050	0.999999	0.999916

Equation (5.108) by minimizing below function:

$$\sum_{j=1}^{M} \left(\tilde{P}_j(T,S) - F(r_{1,j}, \sigma_{1,j}) \right)^2, \tag{5.110}$$

where M is the total number of parallel simulations. The value of underlying bond $P(T,S)$ is determined by:

$$\hat{P}_j(T,S) = F(r_{1,j}, \sigma_{1,j} | \{\hat{a}_k\}), \tag{5.111}$$

where $\{\hat{a}_k\}$ are the least-square estimations of Function (5.110).

testing of the algorithm

We assess the accuracy of such algorithm by comparing the bond price calculated with both the time consuming Monte Carlo simulation mentioned in section 5.4.1 and the LSM method. Table 5.16 presents the ratios of the 1 year zero-coupon bond

Table 5.17: The ratio of 10-year zero-bond price from LSM to Monte Carlo Simulation

	r_0	0.01	0.03	0.05	0.07	0.10
CKLS		0.999575	0.999778	0.999788	0.999676	0.999661
	σ_0	0.01	0.03	0.05	0.07	0.10
SVG	0.0005	1.000000	0.999886	1.000340	1.000370	1.000320
	0.0010	0.999798	0.999711	1.000480	1.001250	1.000120
	0.0020	0.999426	1.000550	0.997592	1.001320	0.998974
	0.0030	1.001230	1.000930	1.000670	0.997450	1.000830
	0.0040	1.000960	1.000160	0.999355	1.003160	1.001650
	σ_0	0.01	0.03	0.05	0.07	0.10
SVJ	0.0005	0.999622	1.000280	1.000610	1.000300	1.000760
	0.0010	1.000530	1.000750	0.998677	0.999174	0.999703
	0.0020	1.000610	0.998099	0.999819	1.001910	1.000850
	0.0030	0.999512	0.999831	0.998466	1.001600	1.000020
	0.0040	0.998640	1.000200	1.001220	0.999569	1.003010
	σ_0	0.01	0.03	0.05	0.07	0.10
SVS	0.0005	0.999543	1.000510	1.000620	0.999470	1.000670
	0.0010	1.001100	0.999354	0.999325	0.999941	1.000280
	0.0020	1.000600	1.000600	0.999054	1.001170	0.999847
	0.0030	0.999073	0.999933	0.998461	0.999604	1.002140
	0.0040	1.000740	0.999454	1.001830	0.996408	1.001760

prices obtained from the Monte Carlo simulation with the LSM to those from the usual Monte Carlo simulation for each candidate models. We examine the difference given various initial values of both r_0 and v_0 which are covering the 95% value range of the corresponding stationary distributions. Table 5.17 lists the results for 10 years zero-coupon bond. For each method, the number of simulations from current time t to T is $100,000$, and further $100,000$ simulations from T to S only for the time consuming Monte Carlo. As we observe from both tables, the results from these two methods are very close for each model with various initial values of r_0 and σ_0. Therefore, we can apply the LSM instead of the time consuming Monte Carlo method.

5.4.3 Implied volatility of the bond option

In this section, we will focus on the effects of latent volatility on option pricing by examining the corresponding implied volatilities. These values are calculated through Black-Scholes formula where the underlying bond price is assumed to follow a lognormal distribution. Recall that the famous formula for European call option on a zero-coupon bond under the $\mathbb{Q}$-measure is:

$$c = P(0,S)N(d_1) - KP(0,T)N(d_2), \tag{5.112}$$

where

$$\begin{aligned} d_1 &= \frac{\ln\left(\frac{P(0,S)}{KP(0,T)}\right) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}. \end{aligned}$$

The implied volatility is the value of σ that equates the formula price (Equation (5.112)) to the market price (here we use the simulated value instead). The nonlinear solver we used here to find σ given c is the *Bisection method*.

As an example, we will consider a series of European options on a 2-year zero-coupon bond ($S = 2$) which can be exercised at the end of month 3, month 6, year 1 and year 1.5. The parameters used in simulation are the estimating results from Chapter

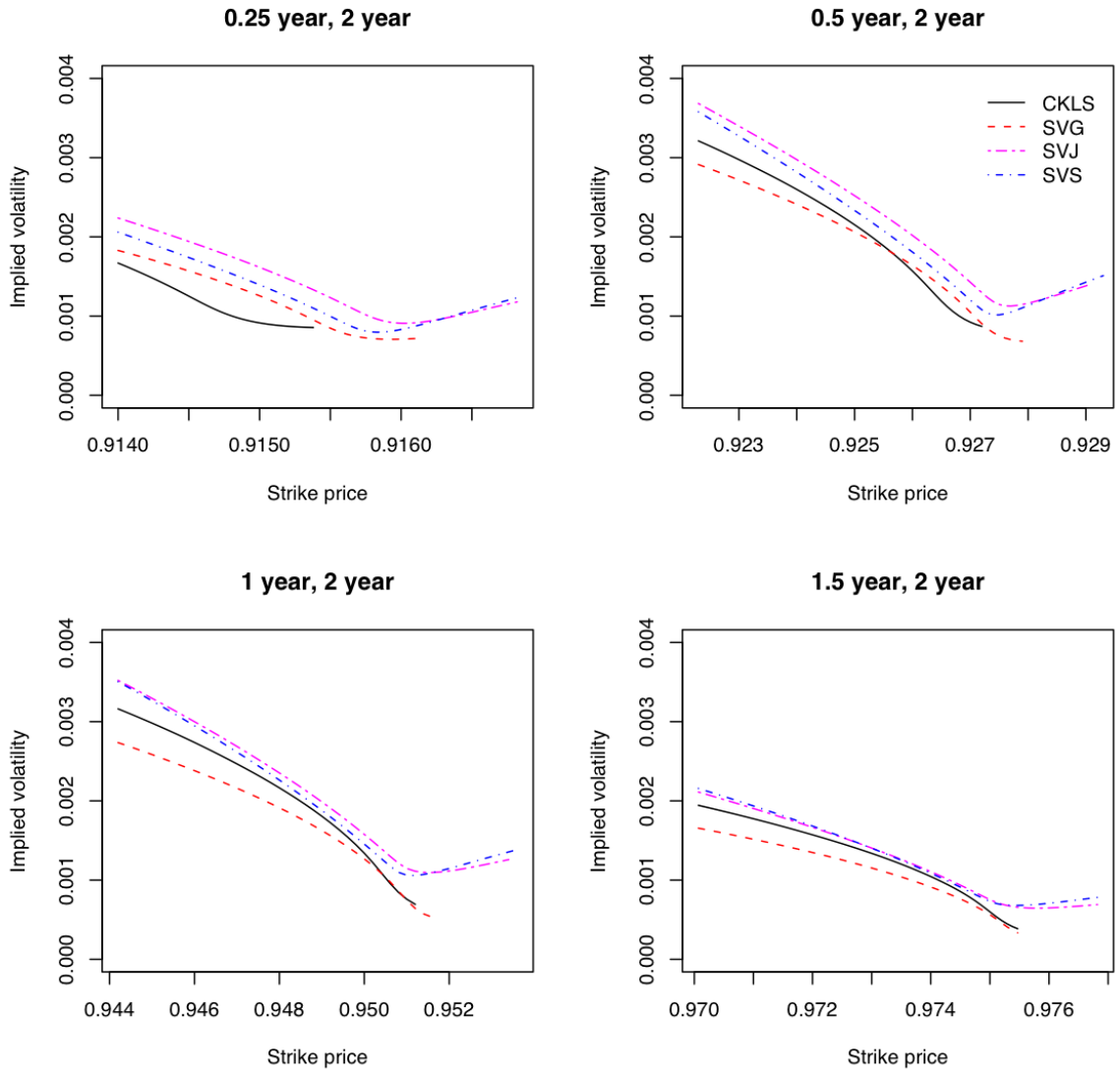

Figure 5.48: Volatility smiles for the CKLS, SVG, SVJ and SVS models. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

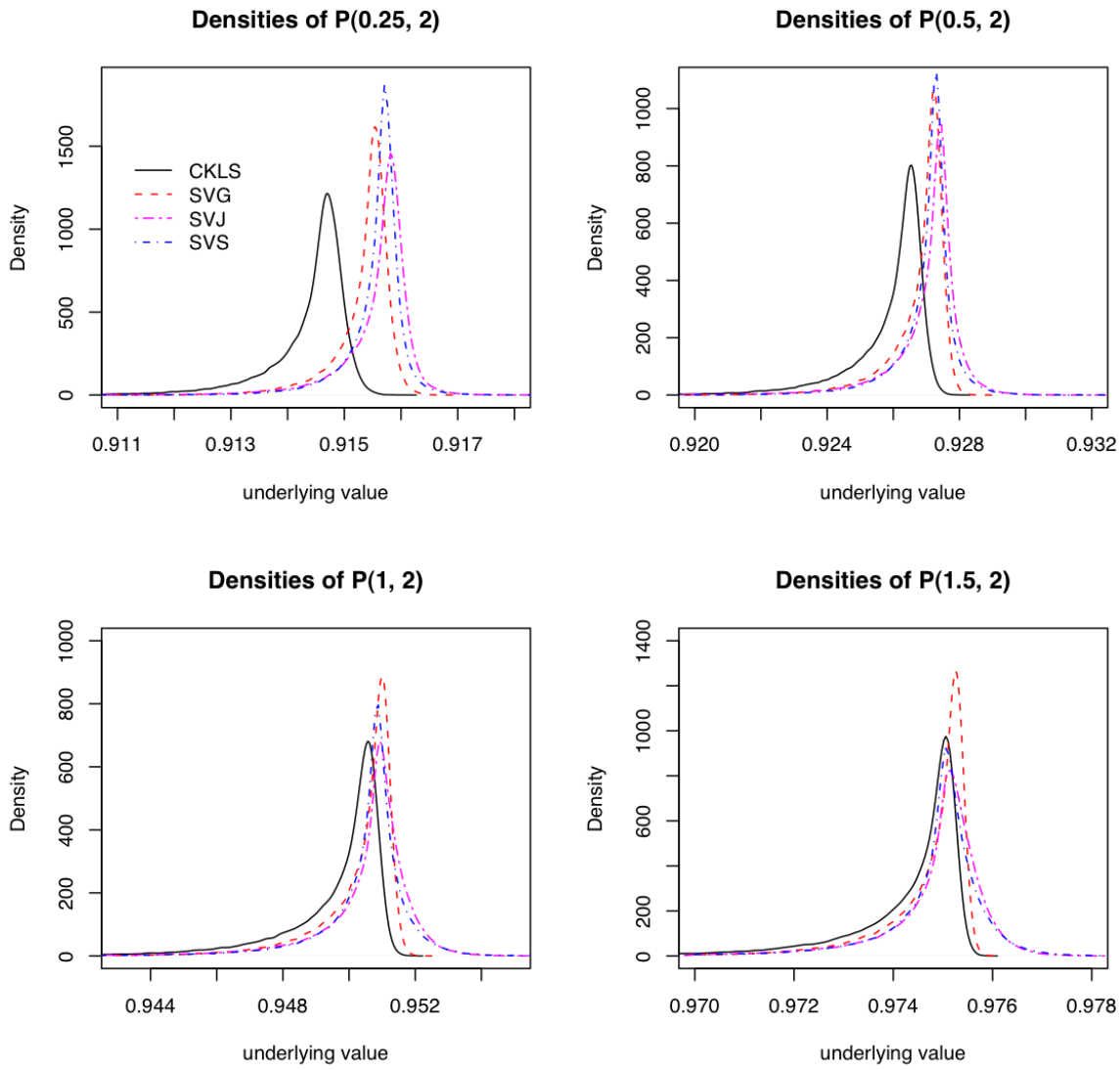

Figure 5.49: Empirical densities of the underlying bond for the CKLS, SVG, SVJ and SVS models. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

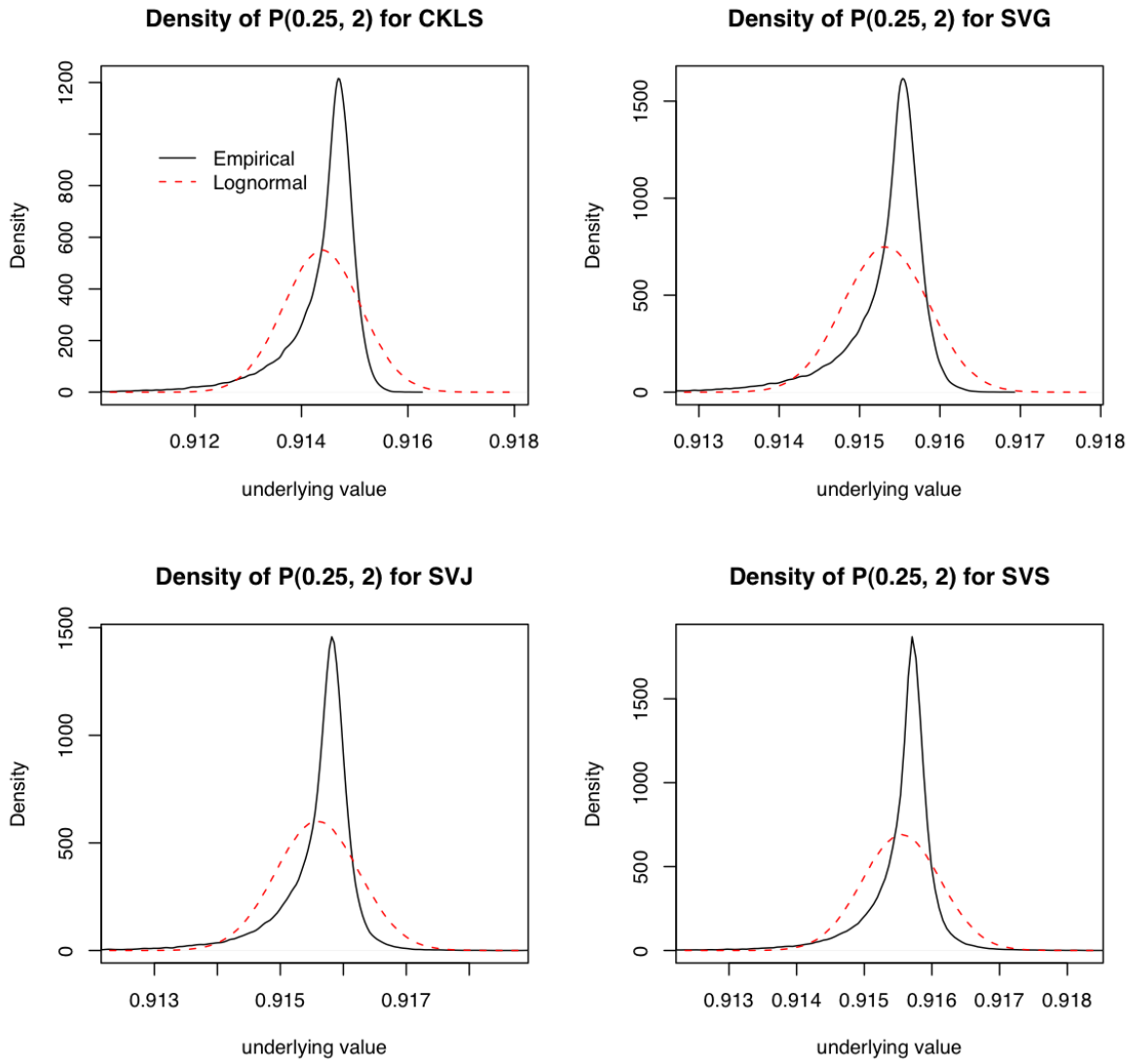

Figure 5.50: The comparison of empirical densities of the underlying bond $P(0.25, 2)$ for various models and the lognormal with same mean and variance. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

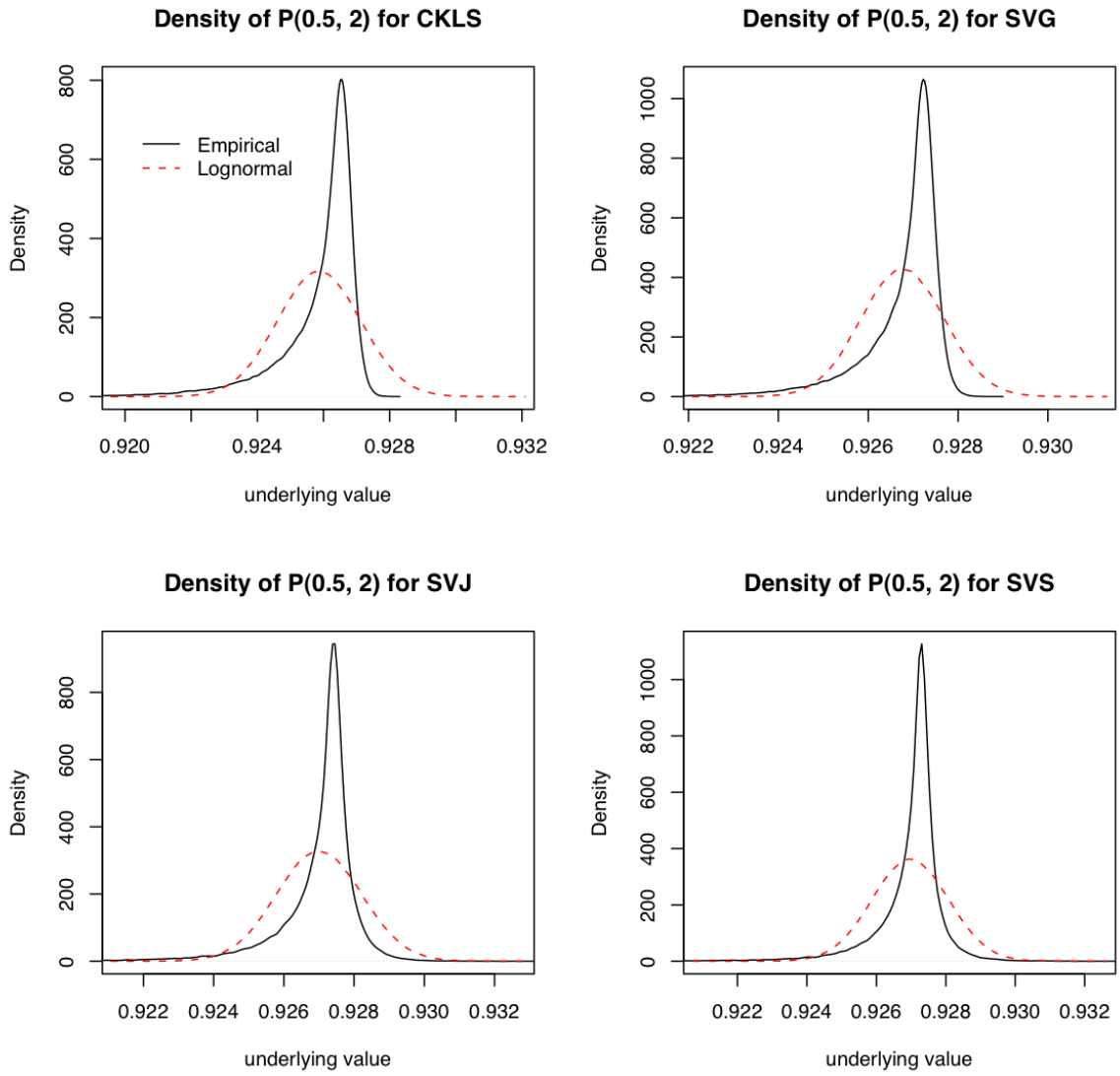

Figure 5.51: The comparison of empirical densities of the underlying bond $P(0.5, 2)$ for various models and the lognormal with same mean and variance. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

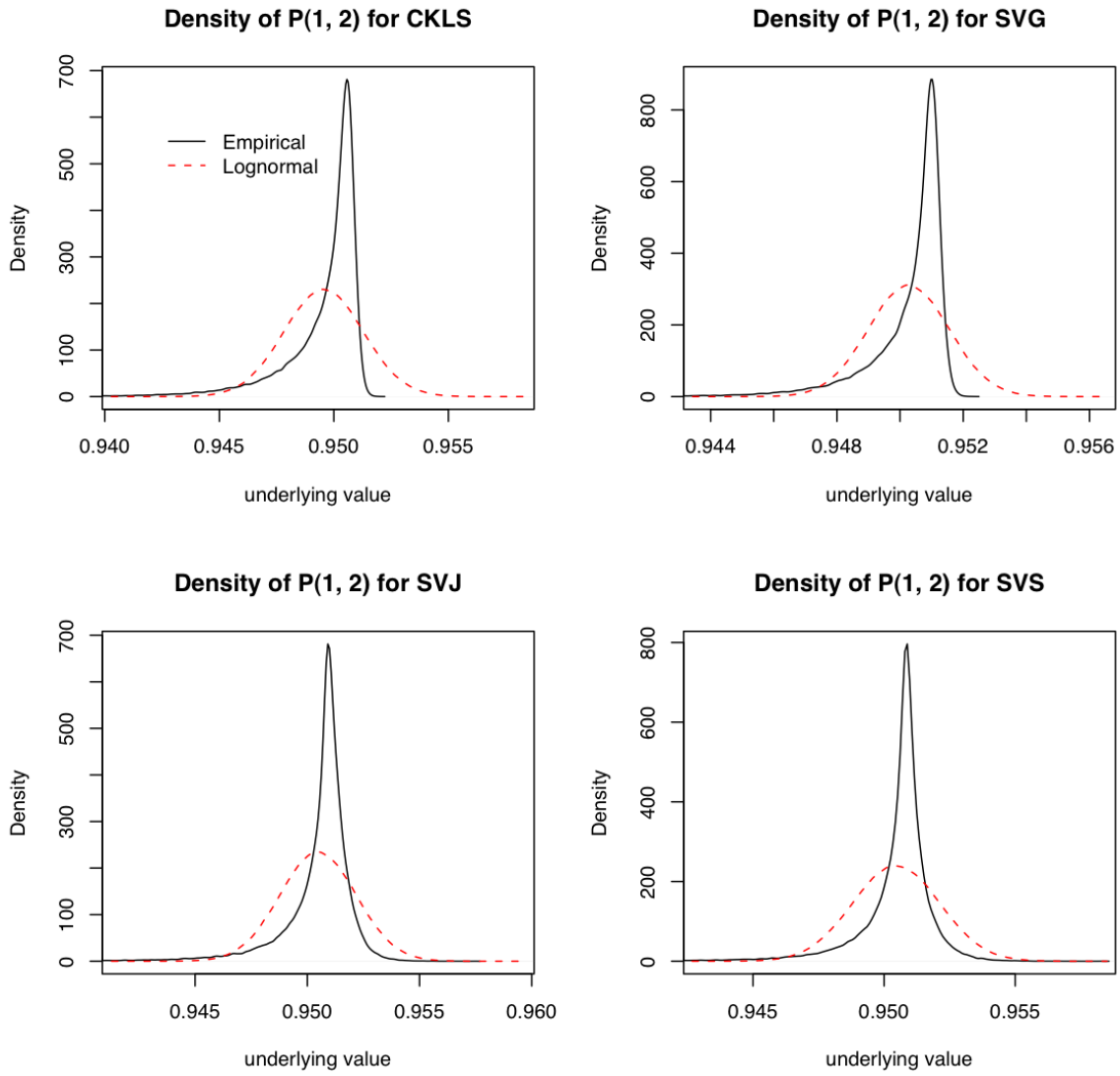

Figure 5.52: The comparison of empirical densities of the underlying bond $P(1,2)$ for various models and the lognormal with same mean and variance. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

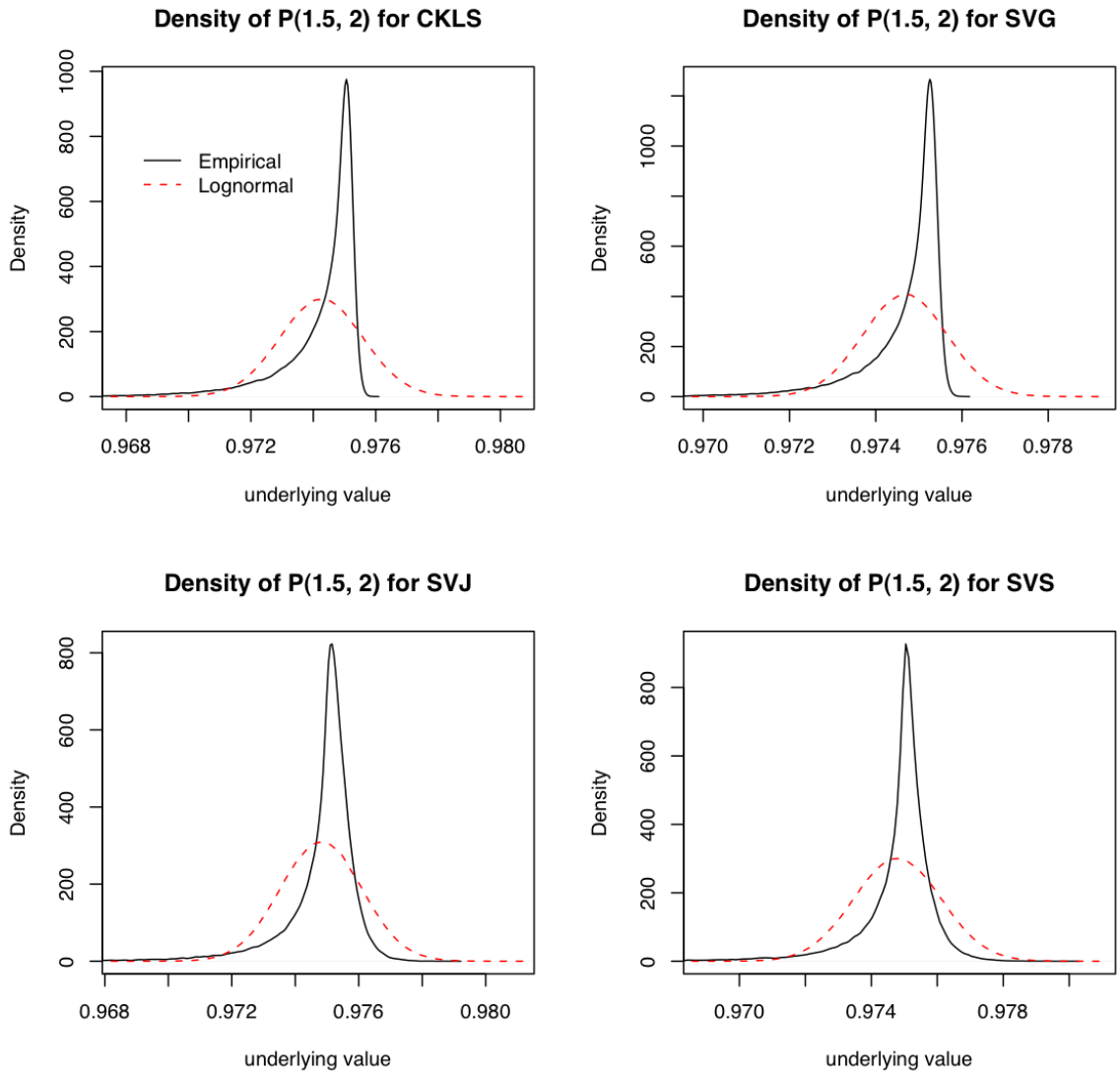

Figure 5.53: The comparison of empirical densities of the underlying bond $P(1.5, 2)$ for various models and the lognormal with same mean and variance. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

4. For each option and every model, the MCMC sample paths are all beginning from the same point. We will also apply the antithetic variate method described before in section 5.3.1 for the interest rate process simulation. The number of simulations for each scenario is 1 million. The implied volatilities are plotted in Figure 5.48, and the empirical densities are presented in Figure 5.49. We also compare these densities with the lognormal distribution with same mean and variance, which are plotted from Figure 5.50 to 5.53.

The introduction of jumps in volatility makes the shape of implied volatility curves quite different as those from the CKLS model and the SVG model. Firstly, the jumps in volatility increase the possible range of the underlying bond price, especially the higher values which can be observed more notably in Figure 5.49. It indicates that the jump process can not only make the volatility to visit high values as the SVG model but also low values while both the SVG model and the CKLS model fail to do so.

Secondly, the jump models have the "hook like" shape right tails which are not present in the CKLS and the SVG model. Eraker *et al.* (2000) also find these kind of shape in the implied volatilities of stock options by introducing jumps in returns and in volatility. The explanation of such shapes can be found by comparing the underlying bond densities from the simulation of candidate models with the lognormal distributions with same mean and variance. Recall, under the Black-Scholes framework, the underlying zero-bond is assumed to follow lognormal distribution. Figure 5.52 presents such comparisons for $P(1,2)$, for example. Because of the low likelihood of visiting higher values, the CKLS and the SVG model only have fatter left tails. However, the jump models have fatter tails on both sides, then the option values from such models are higher than the prices obtained from the Black-Scholes formulas given either low strike prices or high ones.

Thirdly, as the expiry date increases the difference between the SVJ model and the SVS model decreases. The effects of different jumps are more significant in the short

term. Another characteristic worthy of being noticed is higher right tails in the SVS model than it in the SVJ model. This is due to the faster converging jump part in the SVS model which is used to generate the small jumps which happen more often than bigger ones.

Finally, the volatility curves from the SVG model flatten much quicker than the others as the maturities increase. The effect of Gaussian driven volatility process is less significant than the compound Poisson process over the short expiration. Such effect decays also quicker as moves to long time.

As a conclusion, the change of driven process in volatility part has a significant impact in bond option pricing. The jumps in volatility increase not only the visiting range of bond price but also the non-lognormality of the underlying price.

5.5 Caplet

5.5.1 Value of caplet

Consider a caplet on a τ years LIBOR expiring at time $T+\tau$ with a principal of 1, with an annualized strike rate, (*cap rate*), r_c, the payoff of such caplet at $T+\tau$ is:

$$H(T+\tau) = \tau \max(L(T,T,T+\tau) - r_c, 0),$$

where $L(T,T,T+\tau)$ is the LIBOR rate for the period $(T,T+\tau)$ defined by:

$$L(T,T,T+\tau) = \frac{1-P(T,T+\tau)}{\tau P(T,T+\tau)}. \tag{5.113}$$

It is more convenient to value such derivative under the *forward measure* instead of the $\mathbb{Q}$-measure. Under the *forward measure* $\mathbb{P}_{T+\tau}$ where the zero-coupon bond with a maturity of $T+\tau$ is selected as the numeraire, the process of relative payoff value, $\{H'(T+\tau) : H'(T+\tau) = H(T+\tau)/P(T+\tau,T+\tau)\}$ is a martingale. The value of $H(T+\tau)$ at t is then:

$$V(t) = P(t,T+\tau)E_{T+\tau}(H(T+\tau)|\mathcal{F}_t), \tag{5.114}$$

where $E_{T+\tau}()$ is the expectation under $\mathbb{P}_{T+\tau}$-measure.

The Black-Scholes formula for $H(T+\tau)$ can be obtained if the forward LIBOR $L(t,T,T+\tau)$ is a geometric Brownian motion under $\mathbb{P}_{T+\tau}$. Recall that $L(t,T,T+\tau)$ is defined by:

$$L(t,T,T+\tau) = \frac{P(t,T) - P(t,T+\tau)}{\tau P(t,T+\tau)}. \tag{5.115}$$

The SDE of $L(t,T,T+\tau)$ under $\mathbb{P}_{T+\tau}$-measure is:

$$dL(u,T,T+\tau) = L(u,T,T+\tau)\sigma(u)d\hat{W}_u, \tag{5.116}$$

for $t < u < T$, where $\hat{W}_u$ is a Brownian motion under $\mathbb{P}_{T+\tau}$-measure, $\sigma(u)$ is deterministic and $\sigma(u) = 0$ for $T < u < T+\tau$ since $L(T,T,T+\tau)$ is fixed after T. Then $logL(T,T,T+\tau)$ is normal with variance equal to $\int_t^T \sigma(u)^2 du$. The Black-Scholes formula for caplet is:

$$V(t) = P(t,T+\tau)\tau(L(t,T,T+\tau)\Phi(d_1) - k\Phi(d_2)), \tag{5.117}$$

where d_1 and d_2 is defined as:

$$\begin{aligned} d_1 &= \frac{\log(L(t,T,T+\tau)/k) + \frac{1}{2}S^2(T-t)}{S\sqrt{T-t}}, \\ d_2 &= d_1 - S\sqrt{T-t}, \end{aligned}$$

and S is the implied volatility of caplet which equals $\sqrt{\int_t^T \sigma(u)^2 du/(T-t)}$.

5.5.2 Monte Carlo simulations for caplet

Given each short rate model we have considered, the caplet value can be obtained by Monte Carlo simulation. The corresponding implied volatilities are the solutions of Equation 5.117 given the caplet simulation from each model.

The Monte Carlo simulation under $\mathbb{Q}$-measure for the caplet based on Equation (5.114) is

$$\tilde{V}(0) = \frac{1}{M}\sum_{i=1}^{M} e^{-\sum_{j=0}^{N-1}\tilde{r}_{i,j}\Delta t}\tau \max(\tilde{L}_i(T,T,T+\tau) - k, 0)$$

where M is total number of simulations, $N = (T+\tau)/\Delta t$, Δt is one week, $\tilde{r}_j$ is the short rate simulation with $\tilde{r}_1$ the rate at time 0, $\tilde{r}_N$ the rate at $T+\tau$ and the underlying LIBOR is calculated by:

$$\tilde{L}(T,T,T+\tau) = \frac{1-\tilde{P}(T,T+\tau)}{\tau\tilde{P}(T,T+\tau)}$$

with $\tilde{P}(T,T+\tau)$ determined by the Least Square Method (LSM). The other values we needed to apply formula (5.117) are the forward LIBOR $L(0,T,T+\tau)$ over $(T,T+\tau)$, the zero bond $P(0,T)$ and $P(T+\tau)$ which are calculated by:

$$\begin{aligned}
\tilde{L}(0,T,T+\tau) &= \frac{\tilde{P}(0,T)-\tilde{P}(0,T+\tau)}{\tau\tilde{P}(0,T+\tau)} \\
\tilde{P}(0,T) &= \frac{1}{M}\sum_{i=1}^{M} e^{\sum_{j=0}^{N_1-1}\tilde{r}_{i,j}\Delta t} \\
\tilde{P}(0,T+\tau) &= \frac{1}{M}\sum_{i=1}^{M} e^{\sum_{j=0}^{N_2-1}\tilde{r}_{i,j}\Delta t}
\end{aligned}$$

where $N_1 = T/\Delta t$, the sample length for T bond and $N_2 = (T+\tau)/\Delta t$, the sample length for $T+\tau$ bond.

We examine a series of caplets with various pay-time at 6 months, 1 year, 5 years and 10 years, and for each case we also compare the caplets with different tenors: 6 months and 1 year. The number of Monte Carlo simulations is 1 million for each example. The implied volatility curves are plotted in Figure 5.54, Figure 5.56, Figure 5.58 and Figure 5.60 respectively, and the corresponding empirical densities of LIBOR are in 5.55, Figure 5.56, Figure 5.58 and Figure 5.60. To observe clearly of the changing of volatility curves with various maturities, we also plot the curves with half year tenor for each model in Figure 5.62.

There are several differences among the caplet volatility curves of these models. Firstly, the extra random factor in volatility extends the value range of the underlying LIBOR. Furthermore, the range extensions from the jump models are more significant than that from the continuous SV model. It is consistent with the results from bond option. Since the LIBOR is a function of the inverse value of bond, the jump models which enable bond prices to reach higher value help the LIBOR rate

to visit smaller values.

Secondly, the hook-shape left tail is more significant in the jump models over short maturity. As the expiration increases from half year to 10 years, the implied volatility becomes flatter, see Figure 5.62. When we examine the corresponding empirical densities, the fat tails become thinner as time increases especially for the jump models. As pointed in Eraker *et al.* (2000), *the fat-tails and asymmetries in the conditional distribution are driven to a large extent by the volatility process, rather than jumps as maturity increases.* One more point worthy of note is that the curvature of LIBOR with long tenor decays quicker for every case, comparing the volatility plots with same maturity but different tenor (Figure 5.54, Figure 5.56, Figure 5.58 and Figure 5.60).

Thirdly, the difference between the SVJ model and the SVS model is very small over the 1-year maturity case, see Figure 5.56, but a significant difference emerges as maturity rises to 10 years. Since there is only one jump process in the SVJ model used to capture both short term and long term movement in volatility, it is inevitable that the SVJ model will enlarge the effect over short term while reducing it over long term. The SVS models, on the other hand, enable us to capture both short term variation by a jump process with high decay rate and long term in volatility with a smaller one.

In conclusion, the impacts of jumps in volatility to caplet are also very significant as we have seen in the bond options, although the curves for these two kinds of derivatives are different. The difference between the SVJ and the SVS model increases as maturity extends, even though the two models have been fitted to the same historical dataset.

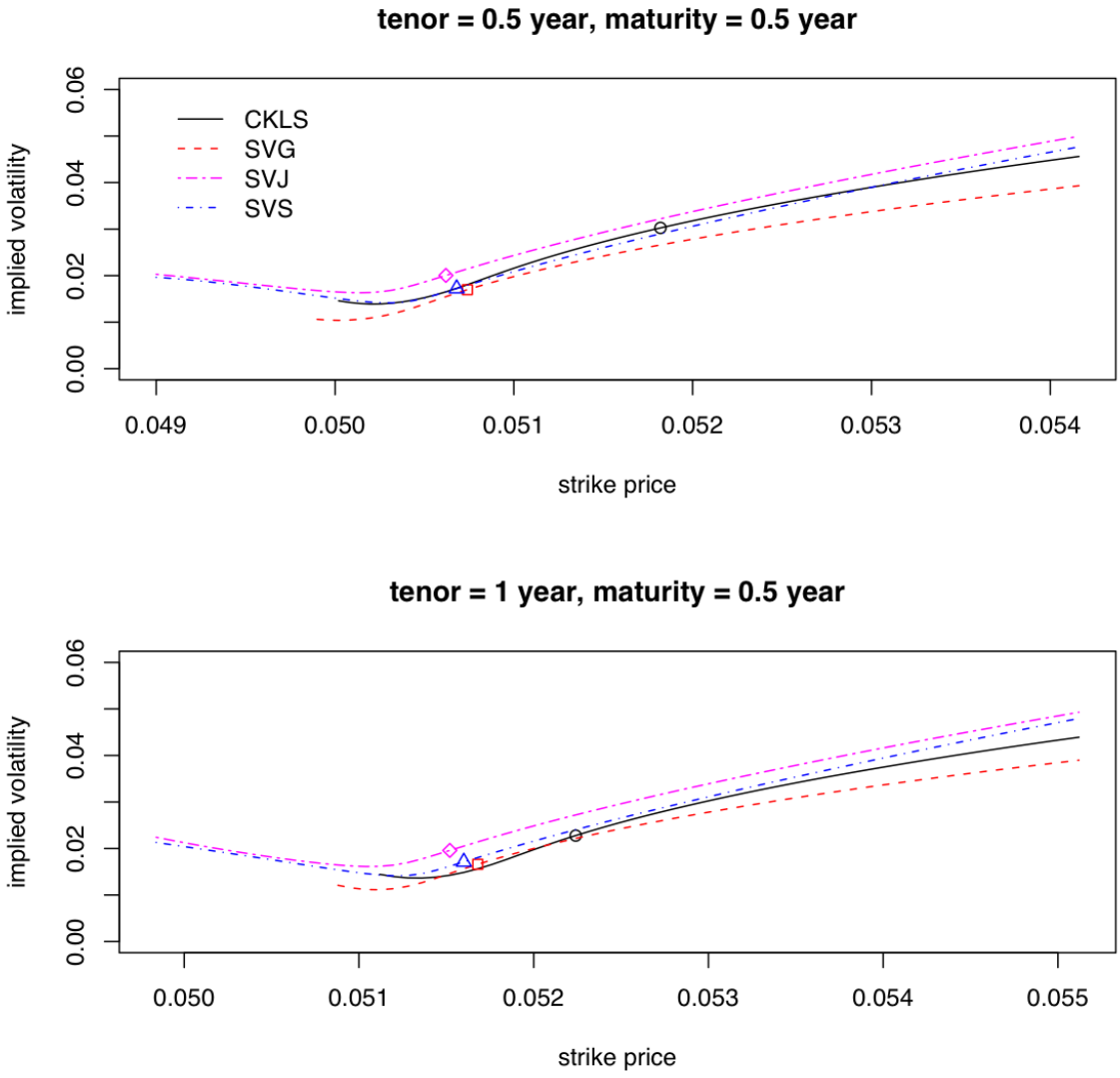

Figure 5.54: Implied volatility curves for the 0.5 year caplet with 6 months tenor and 1 year tenor marked with *at the money* caplet rate. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

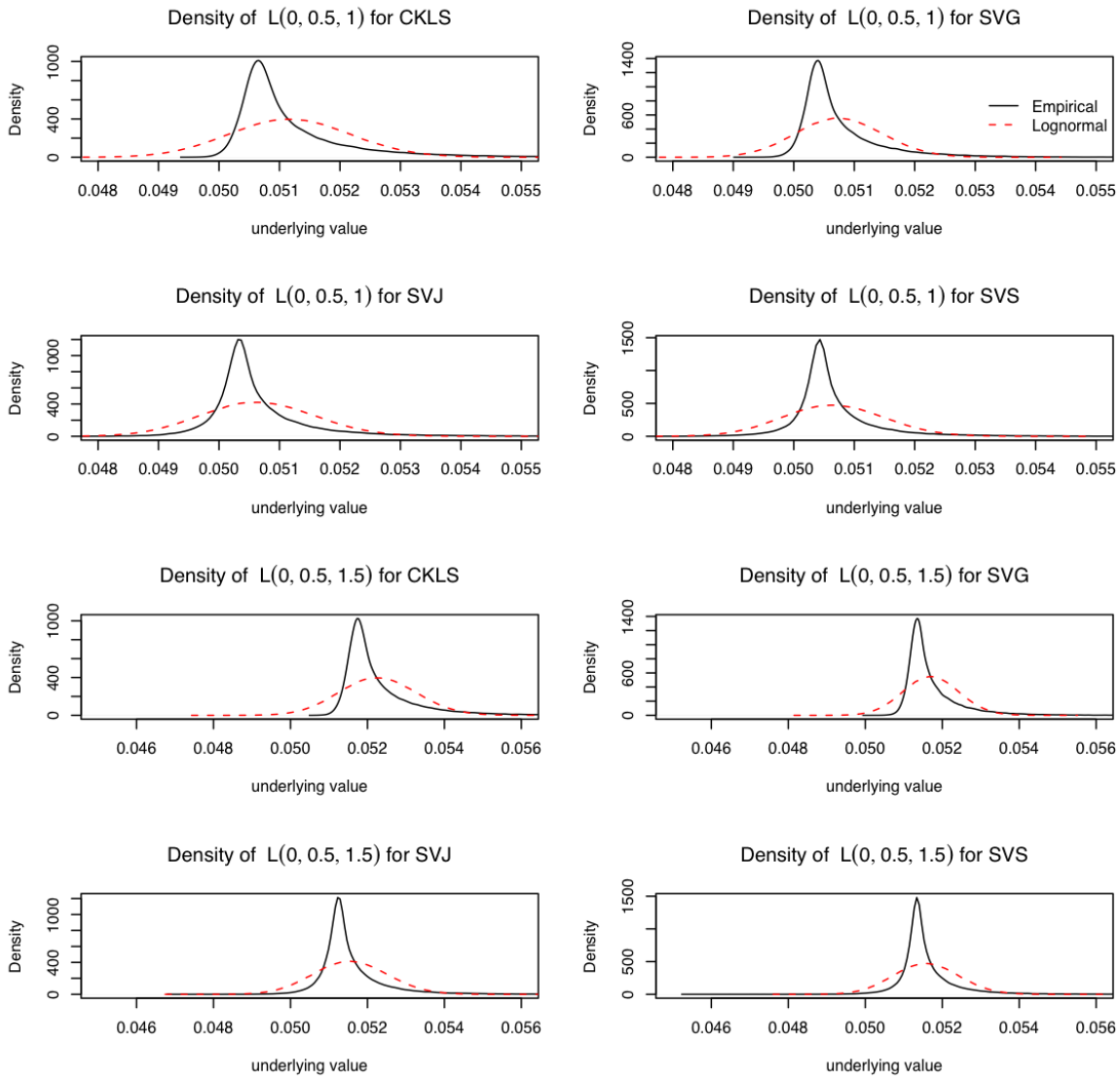

Figure 5.55: Empirical densities of both the underlying LIBOR, $L(0, 0.5, 1)$ and $L(0, 0.5, 1.5)$ for the CKLS, SVG, SVJ and SVS models. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

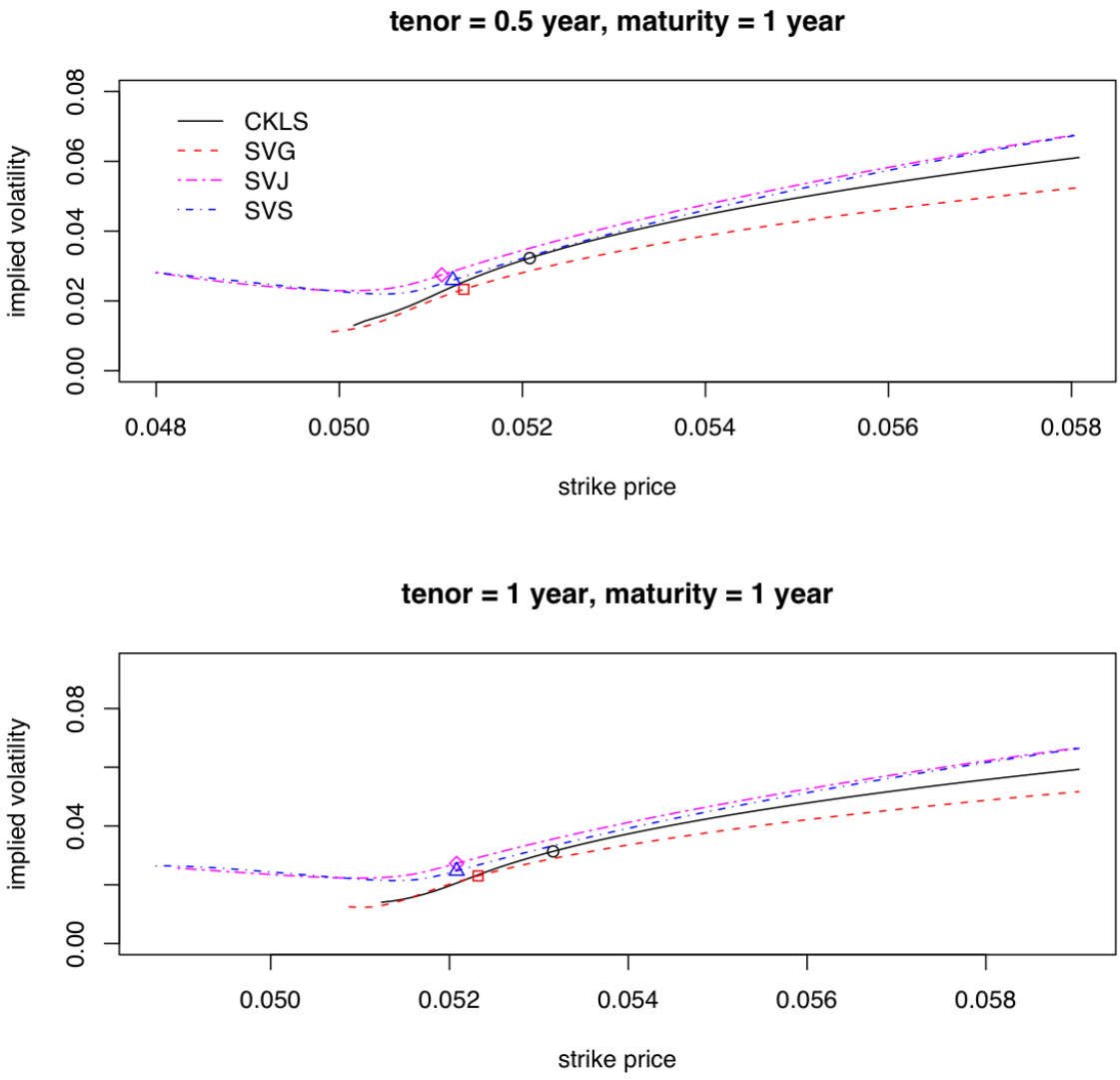

Figure 5.56: Implied volatility curves for the 1 year caplet with 6 months tenor and 1 year tenor marked with *at the money* caplet rate. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

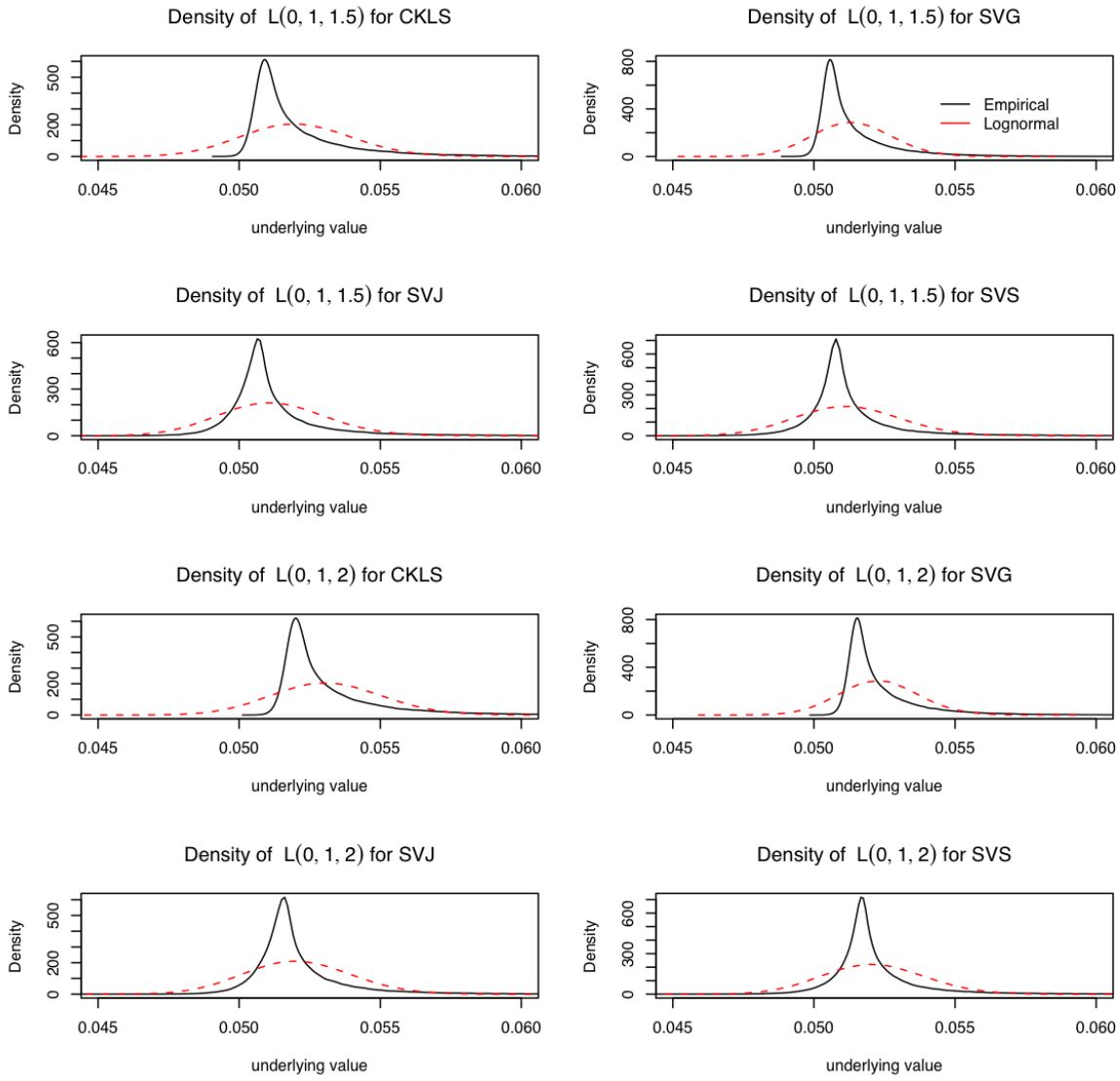

Figure 5.57: Empirical densities of both the underlying LIBOR, $L(0, 1, 1.5)$ and $L(0, 1, 2)$, for the CKLS, SVG, SVJ and SVS models. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

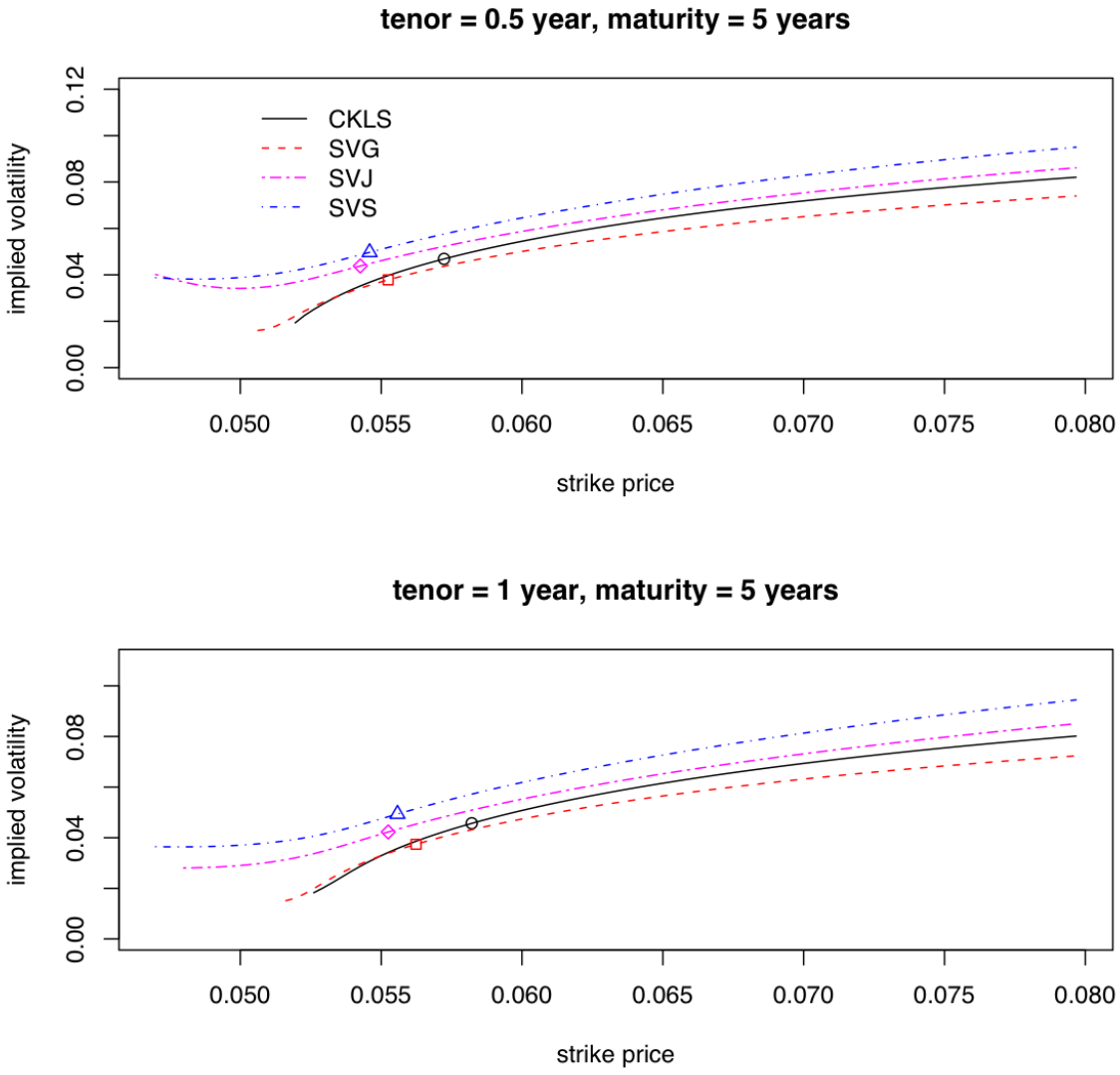

Figure 5.58: Implied volatility curves for the 5 years caplet with 6 months tenor and 1 year tenor marked with *at the money* caplet rate. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

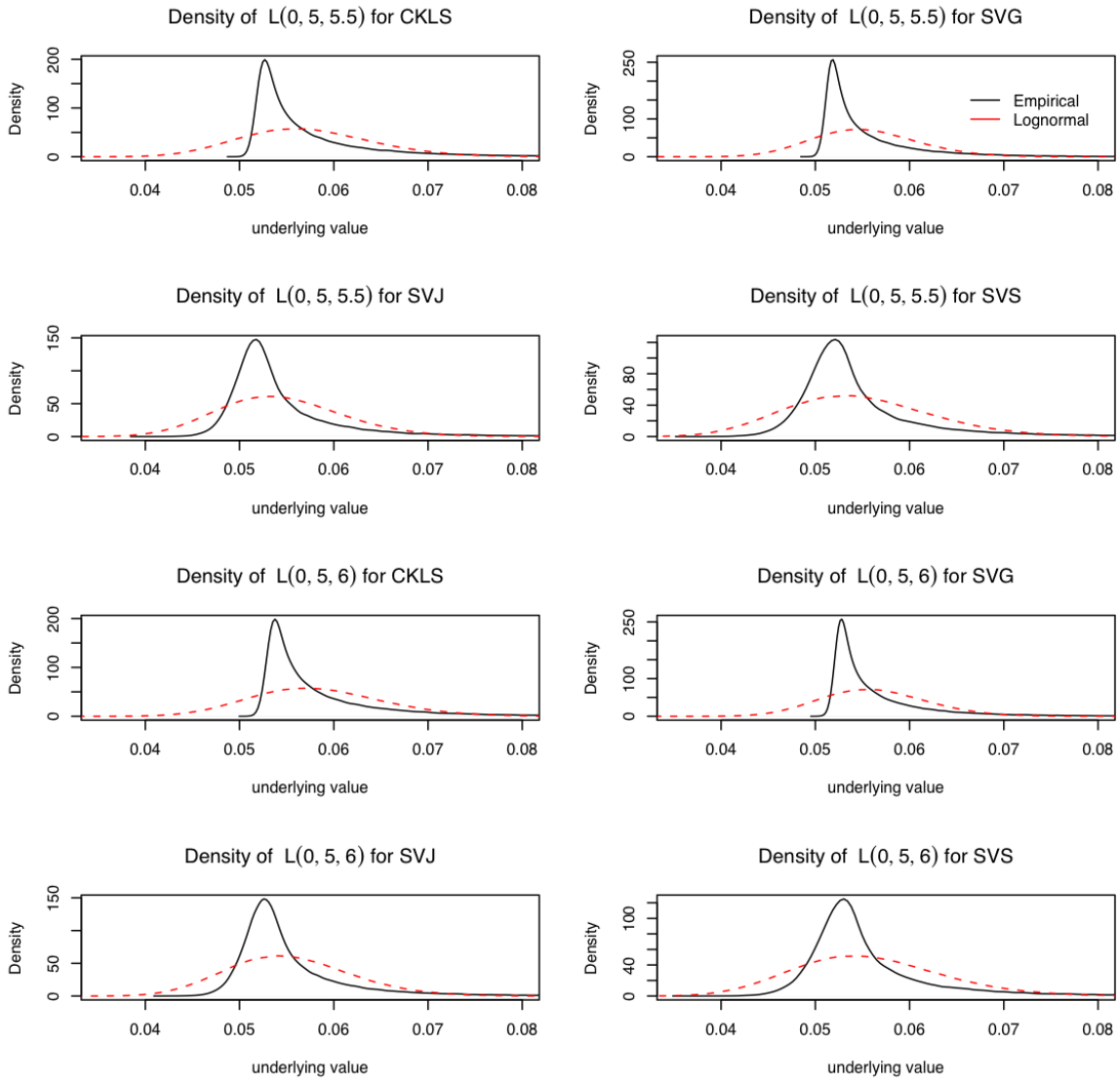

Figure 5.59: Empirical densities of both the underlying LIBOR , $L(0, 5, 5.5)$ and $L(0, 5, 6)$, for the CKLS, SVG, SVJ and SVS models. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

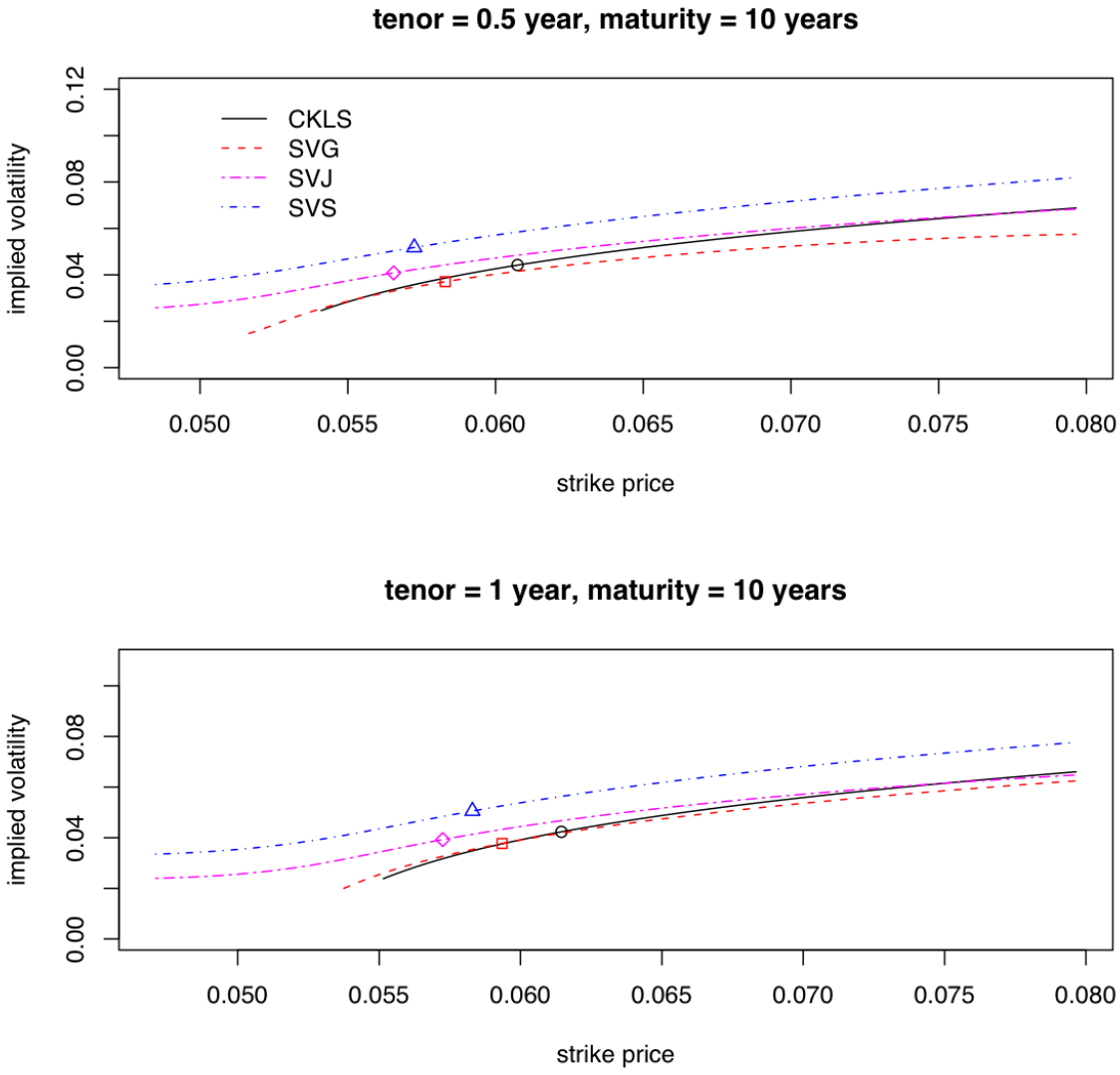

Figure 5.60: Implied volatility curves for the 10 years caplet with 6 months tenor and 1 year tenor marked with *at the money* caplet rate. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

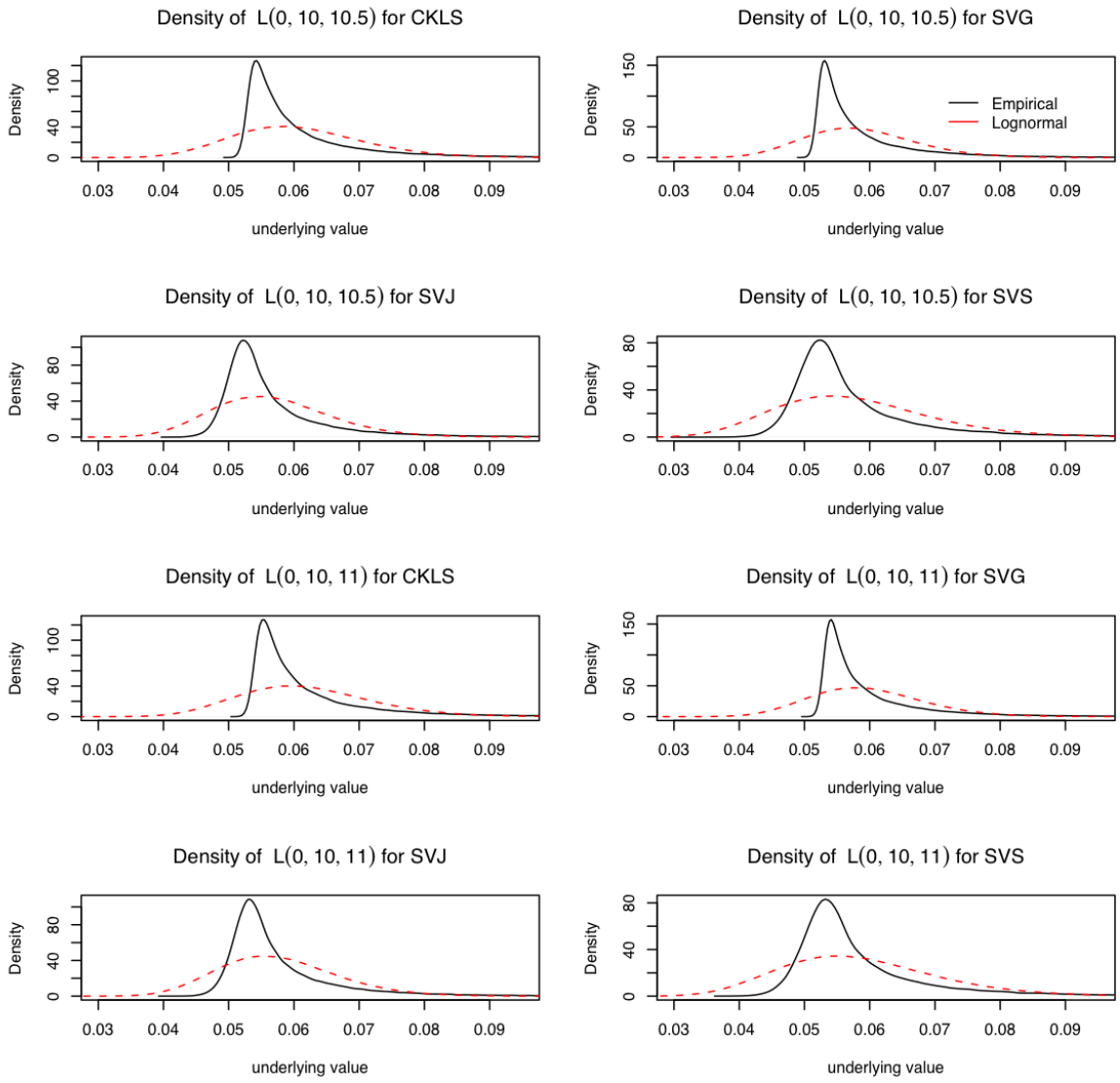

Figure 5.61: Empirical densities of both the underlying LIBOR , $L(0, 10, 10.5)$ and $L(0, 10, 11)$, for the CKLS, SVG, SVJ and SVS models. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

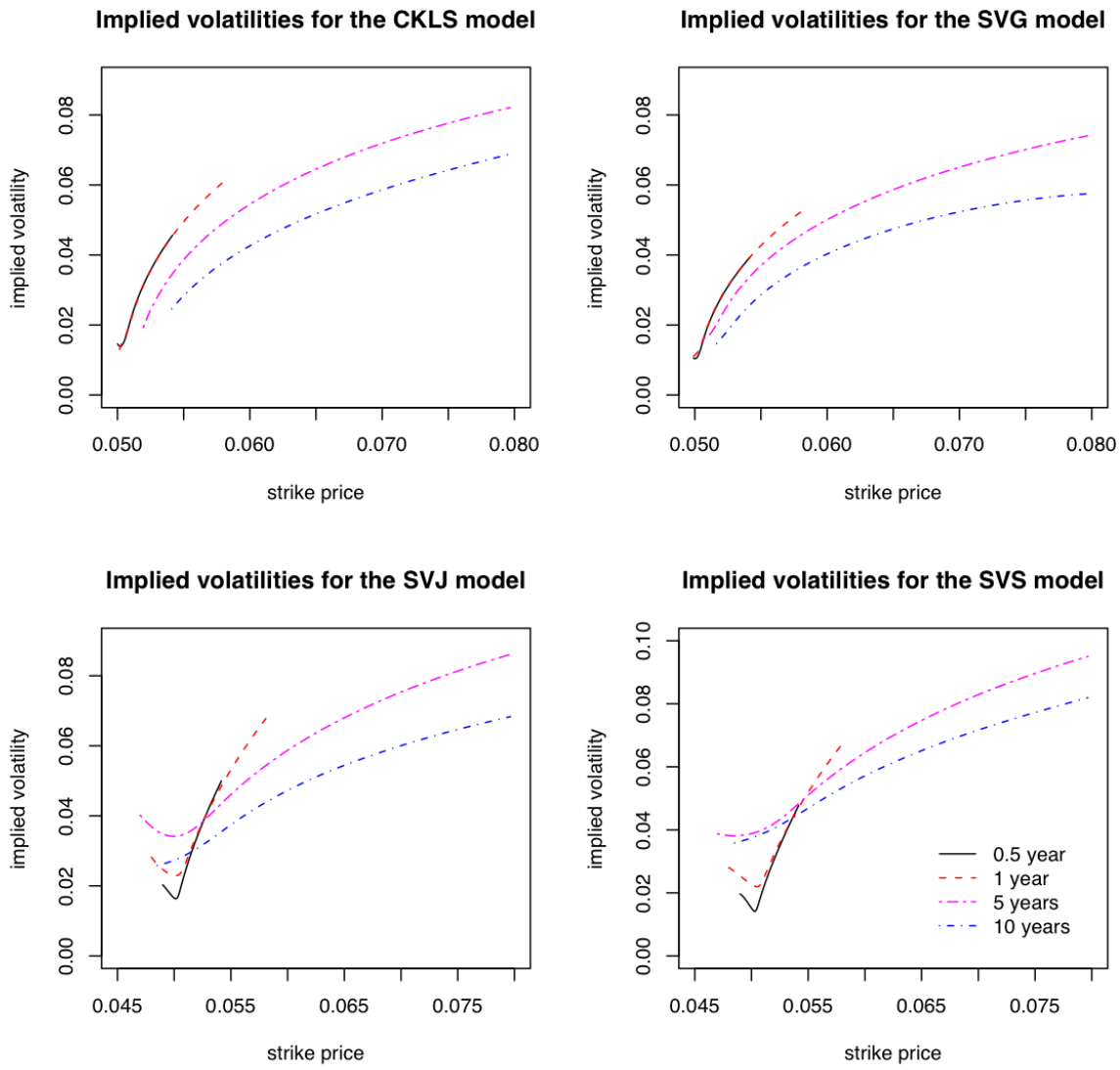

Figure 5.62: Implied volatilities of the caplet with $T = 0.5, 1, 5, 10$ and $\tau = 0.5$. $r_0 = 0.0492$, and $\sigma_0 = 0.0019$.

Chapter 6

Conclusions

6.1 Summary of Thesis

In Chapter 1, we reviewed some definitions of interest rate, common fix-income products, the basic pricing methodologies, interest rate models and the estimation methods for unknown parameters. We define the instantaneous spot rate which is the object of our models, forward rates (both spot and instantaneous one) and LIBOR which will be used later for pricing. The traditional products described are zero-coupon bond, swap, bond option, cap and floor. We will later apply the Monte Carlo simulation to price these derivatives with our models. In order to do that, we introduced the pricing theory as the cornerstone of model application. We start with a general form of single factor model, the CKLS model and the models we studied in later chapters are extension of this model. We also present other kinds of models which based on other interest rates instead of instantaneous short rate. Finally we give a brief description of the General method of moments (GMM), the Maximum Likelihood estimation (MLE) and Markov chain Monte Carlo method (MCMC). The first two are popular and efficient in single factor model while MCMC is an ideal candidate for multifactor models and the one including latent process. We have applied MCMC for all the models we considered.

In Chapter 2, we present the stochastic volatility models we are interested in. First we describe the Lévy process, the compounded Poison process which is used as the driven factor in volatility process. The motivation of introducing new process

in volatility is because of the misspecification Gaussian-driven SV model and the well-known fact that Brownian motion is just a special case of Lévy process. As a criterion, we also reviewed the model with stochastic volatility driven by Gaussian process (the SVG model) and we deduce the stationary distribution of the underlying process in volatility by the infinitesimal generator. We discuss two jump models: the one driven by single compound Poisson process and the one driven by the superposition of two compound Poisson process with different decaying rate. Both stationary distributions of the underlying process can be obtained by checking the characteristic functions which will help us to build the unobservable volatility process for both estimation and simulation.

In Chapter 3, we describe the MCMC algorithms used for the CKLS, the SVG, the SVJ and the SVS model respectively. The simulation test in later section presents a criterion of the performance of each algorithm. For each composition of parameter set of every model, we first figure out the (combine) posterior distribution and then list the procedure of the MCMC algorithm of the model. The Markov chain can be builded up quite strait forward for the CKLS model due to the simple model structure. However it is unable to update the parameters directly for the SV models, since there are the unobserved stochastic volatility process in them. The problem can be solved by the data augmentation method and the latent process is treated similarly as the unknown parameters updated in each MCMC iteration. The MCMC algorithm for the SVG model is inspired by the work of Eraker (2001), and we developed the algorithms for jump models based on the research of Papaspiliopoulos (2003). Later we test these methods by applying them to the corresponding models. The data used are the simulations from each model given the set of predetermined parameter values. Both MCMC sample paths of the parameters and the autocorrelation graphs provide the evidence of the efficiency and credibility of these algorithms.

The estimation results for each model are presented in Chapter 4. The result tables present not only the posterior means and standard deviations of each parameter but also the t-statics which show that all the parameters are significant different

from zero. The residuals are calculated from the historical data given the parameter estimations and stationary latent process. The misspecification of the CKLS model is obvious since it fails both the normality and independence test of the residuals. Improvement is obtained by introducing the extra random factor. The sample paths of residuals for SV model are much closer to the Normal residuals than the CKLS model. Although both the SVG and the SVJ model fails in normality test, the JB statistic for each model is decreasing. The ACF is unable to tell the performance among the SV models, therefore we study their copulas which provide a clear view of the defaults of the SVG and the SVJ model. Considering of the cost of extra factors and parameters, the Bayes factor is an ideal criterion in model selection. The SVS model is superior to the others with the highest marginal log-likelihood. The inabilities of the SVG and the SVJ model can also be observed by investigating the volatility processes.

In Chapter 5, we apply these models to price some fixed-income products, zero-coupon bond, European bond call option and interest rate cap. Since the parameter estimations are obtained from the canonical measure, we first develop the SDE of interest rate under the pricing measure. In simulation we set the market price of risk as zero for all the model, are the parameters used are the results from Chapter 4. By generating the yield curves given various initial values of interest rate and volatility, the shapes from the SV models are much richer than the CKLS model and jump models surpass the SVG model by supporting more types of curves. We apply the least-squares Monte Carlo approach in both bond option pricing and valuing a caplet which is approved to be an efficient simulation method. Another evidence of the superior of the jump models against the SVG and CKLS model are obtained by comparing the implied volatilities for these two products. First the range of underlying is broader in jump models. Second we observed the hook-shaped tails only in jump models, which is consistent with the findings in Eraker *et al.* (2000) but they consider the asset models instead of short rate ones.

6.2 Areas of Further Research

In describing the historical data and pricing, the adventure of changing the underlying driven process in volatility part is quite impressive. However, there is still considerable scope for further investigation of the model and its application. Some possible works for further improvement of the model are listed below.

- Correlations among the random factors

  We assume that the correlation is zero for each SV model we considered. There would be an improvement by adding one more parameter, ρ, the correlation between the driven process in short rate part and the one in volatility part. It would be useful to compare the related SVG model with both independent and correlated jump models.

- Different jump size

  In the SVS model, the jump sizes in both processes are from the exponential distribution with same parameter. A more flexible structure could be obtained if the intensities of the exponential distribution are different.

- Different driven Lévy processes

  Both the SVJ and the SVS model are driven by the compound Poisson process, another simplistic Lévy process as the Wiener process. Whereas other Lévy processes have been widely applied in asset models: variance gamma, Normal inverse Gaussian (NIG), generalized hyperbolic model, and so on. These models can also be used to capture the empirical facts of the short rate.

- Comparing with market price

  In Chapter 5, we only comparing the implied volatility curves among the models. If the market data are available for such products, it would be convincible to make comparison with the real value.

- Market price of risk

  Also in Chapter 5, we applied these models given the pricing measure equal to the real one which indicates the market price of risk is zero. Further work

can examine the application of each model with non-zero market price of risk. As shown in asset model, the implied volatility curves will be affected with different sets of market price of risk.

Appendix A

Lévy-Itô decomposition

Let $\{Z_t\}_{t\geq 0}$ be a Lévy process on $\mathcal{R}^d$ and ν its Lévy measure.

- ν is a random measure on $\mathcal{R}^d \setminus \{0\}$ and satisfies:

$$\int_{|x|\leq 1} |x|^2 \nu(dz) < \infty \int_{|x|\geq 1} \nu(dz) < \infty$$

- The jump measure of Z, denoted by J_Z, is a Poisson random measure on $[0,\infty) \times \mathcal{R}^d$ with intensity measure $\nu(dx)dt$.

- There exist a vector γ and a d-dimensional independent Brownian motion B_t with covariance matrix M such that

$$Z_t = \gamma t + B_t + \int_{z\geq 1, s\in[0,t]} z J_Z(ds \times dz) + \lim_{\varepsilon \to 0} \int_{\varepsilon \leq |z| < 1, s\in[0,t]} z J_Z(ds \times dz)$$

  where $J_Z(ds \times dz) = J_Z(ds \times dz) - \nu(dz)$.

Proof can be found in Cont & Tankov (2004).

Appendix B

Birth & Death Move

We will either add a new point (c, ε) into Ψ with probability q or remove one point (c, ε) from Ψ with probability $1 - q$. The new point is generated by simulating c' uniformly in $(0, t)$ and ε' from $Ex(\theta)$. The removed point is selected uniformly among Ψ. The acceptance probability for the birth move is

$$\alpha(\Psi, \Psi \cup (c, \varepsilon)) = \min(1, p(\Psi, (c, \varepsilon))),$$

and the probability for the death move is

$$\alpha(\Psi \cup (c, \varepsilon), \Psi) = \min(1, p(\Psi, (c, \varepsilon))^{-1}),$$

where

$$p(\Psi, (c, \varepsilon)) = \frac{L(r|\Theta, h_0, \Psi \cup (c, \varepsilon))}{L(r|\Theta, h_0, \Psi)} \frac{\lambda t}{N(t) + 1} \frac{1 - q}{q}.$$

Appendix C

Displacement Move

The point (c_i, ε_i) which will be replaced by a new point (c, ε) is picked up uniformly from Ψ. The strategy to construct (c, ε) is the second method used in Papaspiliopoulos *et al.* (2004). Here c is simulated from a uniform distribution with interval (c_{i-1}, c_{i+1}), and the new jump size $\varepsilon = \varepsilon_i \exp\{-\mu(c - c_i)\}$. Therefore, the acceptable probability for such movement is

$$\alpha(\Psi, \Psi') = \min\left(1, \frac{L(r|\Theta, h_0, \Psi')}{L(r|\Theta, h_0, \Psi)} \exp\{\theta(\varepsilon_i - \varepsilon) - \mu(c - c_i)\}\right),$$

where Ψ' is the replaced point process.

Bibliography

Ahn, C.M. & Thompson, H.E. (1988). Jump-diffusion processes and the term structure of interest rates. *Journal of Finance*, **4**, 52–52.

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate derivative securities. *Review of Financial Studies*, **9**, 385–426.

A.J.McNeil, R.Frey & P.Emerechts (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Andersen, T.G. & Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, **77**, 343–377.

Andersen, T.G., Bollerslev, T., Diebols, F.X. & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, **96**, 42–55.

Attari, M. (1999). Discontinuous interest rate processes: An equilibrium model for bond option prices. *Journal of Financial and Quantitative Analysis*, **34**, 293–322.

Barndorff-Neilsen, O.E. (2001). Superposition of ornstein-uhlenbeck-based type processes. *Theory of Probability and its Applications*, **45**, 175–194.

Barndorff-Neilsen, O.E. & Shephard, N. (2001). Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. R. Statist*, **Soc.B, 63**, 167–241.

Barndorff-Neilsen, O.E. & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Statist*, **Soc.B, 64**, 253–280.

Black, F. (1976). Studies in stock price volatility changes. In *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section*, 177–181, American Statistical Association.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327.

Brace, A., Gatarek, D. & Musiela, M. (1997). The market model of interest rate dynamics. *Mathematical Finance*, **7**, 127–155.

Brennan, M.J. & Schwartz, E.S. (1979). A continuous time approach to the pricing of bonds. *Journal of Banking and Finance*, **3**, 133–155.

Brenner, R.J., Harjes, R.H. & Kroner, K.F. (1996). Another look at models of the short-term interest rate. *Journal of Financial and Quantitative Analysis*, **31**, 85–107.

Brigo, D. & Mercurio, F. (2006). *Interest Rate Models - Theory and Practice*. Springer.

Cairns, A.J.G. (2004). *Interest Rate Models: An Introduction*. Princeton University Press.

Capman, D. & Pearson, N. (2001). Recent advances in estimating term-structure models. *Financial Analysts Journal*, 77–95.

Chan, K., Karolyi, G., Longstaff, F. & Sanders, A. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Monetary Economics*, **39**, 145–172.

Chib, S. & Jeliazkov, I. (2001). Marginal likelihood from the metropolis-hastings output. *Journal of the American Statistical Association*, **96**, 270–281.

Cont, R. & Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman and Hall.

Cox, J.C., Ingersoll, J.E. & Ross, S.A. (1980). A theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.

Das, S.R. (2002). The surprise element:jumps in interest rate. *Journal of Econometrics*, **106**, 27–65.

Das, S.R. & Foresi, S. (1996). Exact solutions for bond and option prices with systematic jump risk. *Review of Derivatives Research*, **1**, 1–24.

D.B.Nelson (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, **59**, 347–370.

de la Vallée Poussin, C.L.X.J. (????). A strong form of the prime number theorem, 19th century.

Dothan, L.U. (1978). On the term structure of interest rates. *Journal of Financial Econometrics*, **6**, 59–69.

Dybvig, P.H. (1995). Bond and option pricing based on the current term structure. In M.A.H. Dempster & S.R. Pliska, eds., *Mathematics of Derivative Securities*, 271–293, Cambridge.

Engle, R.F. (1982). Autoregressive conditional heteroskydasticity with estimates of the variance of united kingdom inflation. *Econometrica*, **50**, 987–1007.

Eraker, B. (2001). Mcmc analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics*, **19**, 177–191.

Eraker, B. (2004). Do stock prices and volatility jump? reconciling evidence from spot and option prices. *Journal of Finance*, **LIX**, 1367–1403.

Eraker, B., Johannes, M. & Polson, N. (2000). The impact of jumps in volatility and returns, working paper, University of Chicago.

Fong, H.G. & Vasicek, O.A. (1991). Fixed income volatility management. *Journal of Portfolio Management*, **Summer**, 41–46.

Geman, S. & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Mathematical Intelligence*, **6**, 721741.

Gilks, W.R., Richardson, S. & Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.

Harrison, J.M. & Kreps, D. (1979). Martingales and stochastic integrals in the theory of continuous trading. *Journal of Economic Theory*, **20**, 381–408.

Hasting, W.K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, **57**, 97–109.

Heath, D., Jarrow, R. & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, **60**, 77–105.

Heston, S. (1993). A closed form solution for options with stochastic volatility with application costs. *Review of Financial Studies*, **6**, 327–343.

Ho, S.Y. & Lee, S.B. (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, **41**, 1001–1029.

Hull, J. & White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, **42**, 281–300.

James, J. & Webber, N. (2000). *Interest Rate Modelling*. Wiley.

Jeffreys, H. (1998). *Theory of Probability*. Oxford University Press.

Litterman, R. & Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, **1**, 54–61.

Litterman, R., Scheinkman, J. & Weiss, L. (1988). Volatility and yield curve, working paper, Goldman Sachs & Co, New York.

Lo, A. (1986). Statistical tests of contingent-claims asset-pricing models - a new methodology. *Journal of Financial Economics*, **17**, 143–173.

Longstaff, F.A. & Schwartz, E.S. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. *Journal of Finance*, **47**, 1259–1282.

Longstaff, F.A. & Schwartz, E.S. (2001). Valuing american options by simulation: A simple least squares approach. *Review of Financial Studies*, **14**, 113–147.

Merton, R.C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, **4**, 141–183.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. (1953). Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, **21**, 1087–1091.

Miltersen, K., Sandmann, K. & Sondermann, D. (1997). Closed form solutions for term structure derivatives with lognormal interest rates. *Journal of Finance*, **52**, 409–430.

Papaspiliopoulos, O. (2003). *Non-centered Parameterisations for Data Augmentation and Hierarchical Models*. Ph.D. thesis, Lancaster University.

Papaspiliopoulos, O., Roberts, G.O. & Dellaportas, P. (2004). Bayesian inference for non-gaussian ornstein-uhlenbeck stochastic volatility processes. *RSSB*, **66**, 369–393.

Pelsser, A. (2000). *Efficient Methods for Valuing Interest Rate Derivatives*. Springer.

Richards, S. (1978). An arbitrage model of the term structure of interest rates. *Journal of Financial Economics*, **6**, 33–57.

Rogers, L.C.G. (1995). Which model for term-structure of interest rate should one use ? *Mathematical Finance*, **65 of IMA. New York:Springer**, 93–116.

Sato, K.I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

Scott, L.O. (1997). Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: Applications of fourier inversion methods. *Math. Finance*, **7**, 413–426.

Stambaugh, R.F. (1988). The information in forward rates: Implications for models of the term structure. *Journal of Finance and Economics*, **21**, 41–70.

Strickland, C. (1996). A comparison of diffusion models of the term structure. *The European Journal of Finance*, **2**, 103–123.

Tanner, M.A. & Wang, W.H. (1996). The calculation of posterior distributions by data augmentation. *J*, **2**, 103–123.

Taylor, S. (1986). *Modelling Financial Time Series*. Wiley.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, 177–188.